

DISSERTATION / DOCTORAL THESIS

TITEL DER DISSERTATION / TITLE OF THE DOCTORAL THESIS

GROUP AGENCY, CONSTITUTIVISM AND RESPONSIBILITY

VERFASST VON / SUBMITTED BY

FRANZ URBAN ALTNER, M.A.

ANGESTREBTER AKADEMISCHER GRAD / IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF:

DOKTOR DER PHILOSOPHIE (DR. PHIL.)

WIEN, 2022/ VIENNA, 2022

Studienkennzahl lt. Studien- UA 792 296
blatt / degree programme code
as it appears on the student
record sheet:

Dissertationsgebiet lt. Studi- Philosophie
enblatt / field of study as it
appears on the student record
sheet:

Betreut von/Supervisor: Univ.-Prof. Mag. Dr. Herlinde Pauer-Studer
Betreut von/Supervisor: Univ.-Prof. Dr. Hans Bernhard Schmid

Contents

Abstract	i
Preface	iii
Introduction	5
1 The Intentional Stance	6
1.1 Introduction	6
1.2 Explanatory Strategies	7
1.3 Basic Elements of the Intentional Stance	8
1.3.1 Rationality	9
1.4 Interpreting an Agent	11
1.5 Irrationality, Explanatory Gaps and the Design Stance	14
1.5.1 Adopting the Design Stance	18
1.6 Ontology and Critical Remarks	20
1.7 The Intentional Stance and Group Agency	21
2 A Thin and Thick Account of Group Agency	23
2.1 Introduction	23
2.2 Collections and Group Agents	24
2.3 Aggregation Functions	25
2.4 The Thin Account of Group Agency	26
2.4.1 Shared Intentions	32
2.4.2 A Closer Look at Aggregation Functions	37
2.4.3 Group Reasoning	41
2.4.4 Group Action Under the Thin Account	44
2.5 The Thick Account of Group Agency	46
2.5.1 Group Identification	46
2.5.2 From Group Identification to Self-Identification	49
2.5.3 The Thick Account of Agency and the Group Will	55
2.5.4 Self-Identification and Collective Reasoning	56

2.5.5	Bratman's Account of Shared Deliberation	57
2.6	Proceduralist Accounts and Supervenience	61
2.7	Summing up	62
3	The Attributionist Challenge	64
3.1	Introduction	64
3.2	Misappropriation and External Ends	64
3.3	The Proceduralist's Response	71
3.4	The Misappropriation Objection and its Metaphysical Consequences	74
3.5	Summing up	77
4	Group Responsibility and the Pathological Group Agent	78
4.1	Introduction	78
4.2	The Reactive Attitudes and the Quality of the Will	78
4.2.1	Excuses	80
4.2.2	Exemptions and the Objective Stance	80
4.2.3	Moral and non-Moral Relationships	83
4.2.4	Two Ways of Holding Responsible	85
4.2.5	List and Pettit's Account of Group Responsibility	87
4.3	Second Challenge: The One-Dimensional Valuer	89
4.3.1	Constitutive Ends and the ODV	91
4.4	The ODV and Responsibility	97
4.4.1	The ODV and the Powers of Self-Control	98
4.5	How do the two Challenges Hang Together?	100
4.6	Summing up	101
5	A Constitutive Account of Group Agency	103
5.1	Introduction	103
5.2	A Constitutivist Theory of Individual Agency	104
5.3	A Constitutive Theory of Group Agency	110
5.3.1	Group Self-intelligibility and the Point of View	111
5.4	Group self-intelligibility	114
5.4.1	The essential sharedness of group self-intelligibility	115
5.5	Self-Intelligibility and a Group's Practical Identity	119
5.5.1	A Diachronic Argument	119
5.5.2	Constitutive Rules and Practical Group Identities	124
5.6	Responding to the two Challenges	129
5.6.1	The Attributionist Challenge	129
5.6.2	Self-Intelligibility and Responsibility	129
5.6.3	Some Metaphysical Remarks	133

5.7	Situating the Constitutive Account	134
5.7.1	Self-Identification vs Self-Intelligibility	135
5.7.2	Pauer-Studer's Constitutive Account	135
5.7.3	Hindricks' Status Account	136
5.8	Summing up	138
6	Group Self-Intelligibility and Corporate Governance	139
6.1	Introduction	139
6.2	The public corporation	141
6.3	The Practice of Shareholder Governance	142
6.3.1	Essentialist Shareholder Governance	142
6.3.2	Deontic Shareholder Governance	149
6.3.3	Anchoring the DSG-rule	156
6.3.4	The DSG-practice as a Practical Identity.	158
6.3.5	Objections against the External Practical Identity . . .	160
6.3.6	Further Remarks on Social Practices	162
6.4	Instrumental Ethics	163
6.4.1	Stakeholder governance	169
6.5	Summing up	172
	Conclusion	176
	Bibliography	177

Abstract

Philosophen*Innen, die zu Gruppenakteuren, wie Unternehmen, Universitäten und Staaten, forschen, haben in den letzten Jahren überzeugende Argumente geliefert, diese Gruppen als rationale und moralische Akteure zu begreifen. Die meisten von ihnen vertreten dabei die Position, dass diese kollektiven Akteure an unserer Praxis der Verantwortungszuschreibungen teilnehmen können. Diese Praxis ist durch die reaktiven Emotionen geprägt, die unterstreichen, wie wichtig uns der gute und schlechte Wille im Zusammenleben mit einem anderen Menschen ist. Es folgt daraus, dass Gruppen, die an dieser Praxis teilnehmen können, auch zu einem Willen fähig sind. Ein zentraler Beitrag dieser Dissertation ist zu zeigen, dass zentrale, funktionalistische Gruppenakteurstheorien, wie die von List und Pettit, keine zufriedenstellende Konzeption eines Gruppenwillens haben. Dafür werden zwei Desiderate für eine adäquate Theorie eines Gruppenwillens entwickelt. Eine solche Theorie muss 1) trennscharf zwischen Gruppenhandlungen, in denen die Gruppe selbst die Handlung lenkt, und Gruppenaktivitäten, in denen die Gruppe keine Kontrolle hat, unterscheiden und 2) erklären, warum gewisse normativ-funktionale Strukturen die Möglichkeit eines autonomen Gruppenwillens unterminieren. In einem dritten, positiven Teil wird eine neuartige Willenskonzeption vorgestellt, die auf der funktionalen Theorie Pettits und Lists aufbaut und diese um Vellemans konstitutivistische Handlungstheorie erweitert. Dabei wird eine kollektive Version des konstitutiven Handlungsziels, Selbst-Intelligibilität, entwickelt. Des Weiteren wird ein Argument vorgestellt, das die Notwendigkeit einer externen, praktischen Gruppenidentität aufzeigt. Eine solche muss als soziale Praxis, die die Gruppengrenzen transzendiert, aufgefasst werden. Dieser Gruppenkonstitutivismus erfüllt nicht nur beide Desiderate, sondern erlaubt auch neue Erkenntnisse zum Einfluss verschiedener Arten von Corporate Governance auf die Möglichkeit autonomen Handelns von Unternehmen. Solche Erkenntnisse sind entscheidend für den moralischen Status von Unternehmen und deren Möglichkeit an unserer bestehenden Praxis der Verantwortungszuschreibungen teilzunehmen.

Philosophers of group agency have argued forcefully that we should view structured groups – with a hierarchy and decision mechanisms – as agents capable of rational and moral agency. These include, but are not limited to, corporations, universities, and states. Most theorists have maintained that moral agency allows groups to participate in our responsibility practice that is shaped by the reactive emotions. Part of this practice is an entrenched care for each other’s quality of will. Consequently, these theorists must also hold that groups are capable of a good or ill will. This dissertation argues that despite these aspirations, most theories of group agency fall short of providing a theory of a group or corporate will. It first surveys List and Pettit’s prominent, functionalist theory of group agency, as well as alternative accounts. Two desiderata for any adequate model of a corporate will are developed. The account must 1) be able to distinguish between group activity in which the group is not in control and autonomous group action and 2) explain why certain normative profiles and functional structures undermine a group’s capacity for moral agency and its capacity to manifest an autonomous will. After arguing that List and Pettit’s theory cannot meet these challenges, a positive proposal is developed that provides a successful answer. It extends List and Pettit’s functionalist picture with a collective version of David Velleman’s constitutive aim of agency: Self-Intelligibility. As part of this proposal, a novel way to capture a group’s practical identity is developed that builds on the notion of constitutive rules to model social practices. The resulting constitutive theory of group agency is able to satisfy both desiderata and provides insights into how different types of corporate governance of public corporations undermine and enable their capacity for moral and autonomous agency.

Preface

My fascination with philosophy of action began during my undergraduate studies. Not only is it concerned with questions intersecting with numerous other philosophical fields, a reasonable account of action is also hugely important for non-philosophical disciplines such as economics, law, psychology and artificial intelligence. One of the claims I defend in the dissertation is that existing accounts of group will are ill-suited to address collective versions of challenges well known to trouble their individualist counterpart. Although the idea of a corporate will faces challenges familiar from the philosophy of action, the solutions that we can appeal to are not easily transferable. This dissertation is a step towards an account of a corporate will that might, on the one hand, clarify debates about corporate governance and questions of corporate responsibility, but, on the other, also affords us a better understanding of the relationship between the individual agent as a center of agential control and a group agent. As corporations grow more complex and powerful, and as more corporate tasks are automatised or replaced by algorithms, there is surely more work to be done.

My work on this dissertation has been facilitated by the European Research Council project “The moral and normative foundations of group agency” (grant number 740922) led by Herlinde Pauer-Studer, as well as the Institute of Philosophy at the University of Vienna. In my quest to understand the nature of corporate action and the moral status of group agents such as corporations, I have learned and benefited enormously from numerous conversations and written comments. First and foremost, I want to thank my two supervisors, Herlinde Pauer-Studer and Hans Bernhard Schmid. I have benefited much from their philosophical insights. Those insights have informed central intuitions that have shaped the resulting account of group action. I owe Herlinde Pauer-Studer a special debt of gratitude for her support of my development as a member of the philosophical community: I would not have learned about the practical dimensions of the profession without her guidance.

I would also like to thank the members of the Jacson Group on social

ontology, to whom I presented most drafts of the chapters in this dissertation and whose input was a huge help. I also want to thank the participants of the reading group on David Velleman's book *How We Get Along*, where I learned much more about Velleman's constitutive theory than I could have ever done on my own. Special thanks also goes to the discussion groups within the ERC-project: these sessions have been one of the most rewarding philosophical experiences during my PhD and have informed and improved many of the arguments in this dissertation.

I am further indebted to many others for helpful conversations and comments concerning one or more drafts of the papers in this dissertation. In alphabetical order, thanks to: Gunnar Björnsson, Michael Bratman, John Broome, Stephanie Collins, Felix Darnowski, Niels de Haan, Frank Hindriks, Conrad Josef, Niklas Kirchner, Olof Leffler, Judith Martens, Carlos Núñez, Grace Patterson, Franziska Poprawe, Matthew Rachar, Vitor Sommariva and Leo Townsend.

I have also received helpful comments from audiences and participants at the following conferences: the 2022 Zicklin Center Workshop in Normative Business Ethics at the University of Pennsylvania, the 2021 online MANCEPT Workshop "The Normative Profile of Collective Agents", the 2020 MANCEPT Workshop "Law, Agency and Collective Lives", Social Ontology 2020 - The 12th Biennial Collective Intentionality Conference, the 2020 Workshop "Group Agency and Collective Responsibility" in Gothenburg, "Ethics and Capitalism" – the 26th International Vincentian Business Ethics Conference (IVBEC) in Dublin, the Social Ontology 2019 - 6th Biennial ENSO Conference in Tampere, the Summer School "Public Philosophy and Social Ontology: Rights and Responsibility for a Just Society" in Groningen, the 2019 Workshop "Group Agency and Collective Responsibility" in Vienna and the 2018 Workshop "Morality and Rationality Revisited" in Leeds.

I want to thank Triinu Eesmaa, Niels de Haan, Moritz Altner, Heike Altner, Joerg Altner, Larissa Schlicht and Stefan Funcke for their friendship, support and care during the writing of this thesis.

Finally, I owe a different kind of gratitude to my wife Sarra for her forbearance and support during the long nights it sometimes took me to write this dissertation. As my work was interrupted by a serious injury in the finishing stages, I remain forever thankful to her, my family and friends for helping me through this difficult time.

Introduction

A few years ago, Samsung issued a recall for the Galaxy Note 7 only a few weeks after the product's launch. As a response, former CEO Gregory Lee signed a letter, which was published in three major US newspapers, the *Wall Street Journal*, *New York Times* and *Washington Post*. Here is a short excerpt from that letter.

An important tenet of our mission is to offer best-in-class safety and quality. Recently, we fell short on this promise. For this we are truly sorry. We take our responsibility serious to address concerns about safety and quality. In collaboration with government agencies and industry partners, we are taking pro-active steps to do better.¹

This short quote makes some interesting assertions. It is formulated with the *first-person plural pronoun*. It refers to Samsung's *mission*, its *promises* to consumers, and how those promises weren't kept. Furthermore, it mentions Samsung's *responsibility* to collaborate with agencies and its plans for future reparations. If it had been formulated with the first-person singular pronoun, these attributions of plain moral agency would be nothing special. But, at the same time, they wouldn't quite make sense in this case. Gregory Lee is not directly responsible for the design, and might have never been involved in it. Nor will he himself participate in future changes in production, or personally collaborate with other government agencies. Even if he himself was 'truly sorry', this means little if this doesn't in effect the corporation's response. Similarly, as lawsuits were filed, they were not filed against Gregory Lee, but against Samsung as an independent agent.

This language and its implications raise the question: what kind of agents are corporations like Samsung? Judging from the wording of Lee's letter, they seem to be moral agents, capable of planning, falling short of expectations, and being held responsible.²

¹*Samsung runs full-page apology ads over Galaxy Note 7 recall* (n.d.)

²Some philosophers believe this is a rhetorical trick, by which Gregory Lee intends to

Philosophers who take a functionalist approach towards agency have come up with several sophisticated arguments that show why structured groups – groups with a hierarchy and established decision mechanisms – should be viewed as agents, capable of rational action.³ Moreover, most proponents of functionalism about structured groups argue that these groups are not only capable of rational, but also of moral agency: they may be held responsible for falling short of our moral expectations, and praised for exceeding them. In these debates, responsibility is often understood via Strawson’s (2008) famous insight, that our practice of responsibility is characterized by the reactive attitudes. Strawson has argued that the function of reactive attitudes, such as guilt, indignation and resentment, is to express how much we care about the quality of each other’s will. Therefore, these theorists are also committed to the claim that corporations are capable of acting with an ill or a good will, for which we can, in turn, hold them responsible.

Given these claims, it might come as a surprise that these theorists, with a few exceptions,⁴ adopt a minimalist, Davidsonian model of action. That model states that a person acts when their behavior is characteristically caused by reasons consisting of a desire-belief pair. Albeit this is an appropriate model of motivated activity, as the debate in the philosophy of action in recent decades has shown, it does not capture the volitional and reasoning-centered aspect of the will. It is this aspect that we care about in our practice of responsibility.

In this dissertation, I argue that this gives rise to two challenges that any account of a group will must meet. First, it must distinguish group activity in which the group agent is not in control from self-governed or autonomous group action in which the group, as an independent locus of agency, is in control. This question is, in part, about how to attribute members’ actions to groups. As I show in the dissertation, this problem becomes especially acute in the case of rogue group members who hijack the group to further their own interests.

deflect responsibility away from himself or individual members of the corporation onto an abstract entity which has, as Coffee (1981) has aptly put it, ‘no body to be kicked and no soul to be punished’. On this view, talk of corporations taking responsibility, planning, and acting is a mere *façon de parler*. Although occasionally highly complex and difficult, such talk may be reduced to referring to the actions and interrelations of individual group members, such as contracts and agreements between them. This is known as the nexus of contracts conception of the firm. It’s one of the main approaches to analysing the firm in economic and management literature. It is defended philosophically by Ludwig (2017a) and Bratman (2017).

³List & Pettit (2011), Björnsson & Hess (2017), Hess (2012) and Hindriks (2008)

⁴Philip Pettit’s account of group self-identification is such a theory that I discuss extensively in chapter 2 and 3.

The second challenge relates to the fact that certain groups are set up in a way that is conducive to their fulfilling the requirements of rationality, but not the characteristic requirements for moral agency. Since many theories of group agency equate rational group agency with moral group agency, pathological group agents that meet the conditions of the former, but not the latter, present a problem. Ideally, a theory of a corporate will is sensitive to such structural deficiencies and able to show how they undermine the group's ability to act autonomously.

These challenges present two novel objections to which most theories of group agency have no answer. I argue that these failures are, in a certain sense, connected. They can be traced back to a feature of group agents that is often missed in the current debate on group agency. Groups like corporations, universities, football clubs, churches, NGO's and states are, like the Roman god Janus, double-faced. One aspect has already been mentioned. In principle, group agents are capable of *moral agency*, that is, the capacity to respond to moral reasons.

But this is not the only way that corporations and other groups are viewed. Presented under a different light, they are *tool-like*: brought into existence and designed to serve a pre-determined purpose, and eliminated, redesigned or split into many parts if they don't. Given this instrumental aspect, their intentional activities, though purposeful, seem to lose the name of action. The literature on group agency has, to date, neglected this metaphysical feature impacting the moral agency of group agents. This contrasts with more applied disciplines such as business ethics, where the debate on corporate governance can be seen as an attempt at reconciling this instrumental face with the various moral demands we hold group corporations to.

The central thesis of this dissertation is that there is an available model of a corporate will which can successfully meet both challenges as well as account for the instrumental face of groups. This account builds on the functionalist framework for group agency developed by List & Pettit (2011). It extends that framework with a collective version of Velleman's (2006, 2009b, 2009a) constitutive aim of autonomous human agency: self-intelligibility. A significant way in which the notion of group self-intelligibility diverges from its individual counterpart is that the group agent's practical identity is interpreted in an external way, using the notion of constitutive rules explored and analysed by Searle (2005, 2006). The resulting *constitutive theory of group agency* will allow me to both identify the conditions ensuring self-governance and autonomous action by group agents like corporations, states, churches and universities, as well as the conditions under which they fall short of this aim. It also provides new insights about how different forms of corporate governance impact a group's ability to participate in a practice of responsibility

shaped by reactive attitudes.

Here's how this dissertation will pursue this aim. In my first chapter, I set out the broader methodological context within which I will develop my account of group agency. I do this by investigating in detail Dennett's intentional stance, which underlies both List and Pettit's theory of group agency and Velleman's theory of individual agency. In this chapter, I aim to lay out the space of metaphysical possibilities for a system's agency afforded to us by the intentional stance. The tools that are developed there will allow us to, later on, question and criticize List and Pettit's assumption of supervenience. That assumption, as I argue later on, lies at the heart of some of the problems which List and Pettit encounter in developing a plausible account of group will.

In the second chapter, I explore List and Pettit's account of group agency in depth. It can be divided into a thin and a thick account. The authors employ the former most often in later papers, and are typically associated with it. Philip Pettit alone endorses the latter. The thick account of group agency introduces the notion of group identification and group self-identification. The aim of this chapter is to introduce and investigate the most important building blocks of List and Pettit's theory of group agency, including their notion of aggregation function, incentive compatibility, group identification and group self-identification. I situate these concepts in the general social ontology literature and analyse the notion of group action or group will that underlies both accounts. I also investigate a plausible extension of their theory built on Bratman's account of shared deliberation, which I take to be sufficient for robust moral group agency and which will be my target of two challenges discussed in the subsequent chapters.

In the third chapter, I set forth my first challenge for accounts of group wills. Most accounts, I argue, fail to distinguish between, on the one hand, belief-desire pairs which give rise to intentions that speak for the agent, and, on the other, those which are external to the group and do not count as reasons constituting intentional group actions. After considering several strategies whereby accounts like List and Pettit's might fend off this challenge, I conclude that their problem in drawing this distinction results from their metaphysical commitments.

In the fourth chapter, after expounding on the Strawsonian notion of responsibility that most accounts of moral group agency rely on, I develop the second challenge. To do this, I first specify the interpretation of two distinct ways of holding responsible, responsibility as accountability – corresponding to the capacity to engage in moral argument – and aretaic responsibility – referring to being held responsible for one's character. Next, I construct the normative profile of a specific group agent with a constitutive, non-moral

maximizing aim. That agent, I argue, is a one-dimensional valuer: it always resolves conflicts among competing values with an eye to a single, fixed value. One-dimensional valuers present a problem for List and Pettit's theory of group responsibility, because they are incorrectly identified by them as agents who can be held accountability responsible.

In the fifth chapter, I present a constitutive account of group agency as a solution to the two challenges raised earlier. I begin by introducing Velleman's individual constitutivist account. Velleman argues that the aim of self-intelligibility is constitutive of autonomous actions. Developing a corresponding notion of group self-intelligibility, I argue that this notion entails aiming at making sense of actions from the point of view of the group. Within a structured group agent, such sense-making is essentially a shared endeavour, often involving a practical identity that is put in place not by the group, but by externally anchored social practices which need not involve the group's members. Next, I discuss how understanding autonomous, moral agency in this sense can help us to address the two challenges I raised in the previous chapters, and how the constitutive account relates to existing accounts.

In the final chapter, I consider how we may apply my constitutive account to a debate in business ethics regarding the end of corporate governance and the consequences of adopting different corporate aims or functions for public corporations' capacity for autonomous actions. Specifically, I investigate whether and under what conditions a public corporation may count as a fully-fledged moral agent, with whom we may see eye to eye in moral relationships. I conclude that, under certain social and economic circumstances, such relationships are impossible and that we should manage and control those public corporations which occupy these social niches. This does not mean that some reactive attitudes towards these groups, such as hate and anger, are inappropriate. Rather, it means only that the core reactive attitudes which are characteristic of moral relationships between individual human agents, such as guilt and indignation, cannot exhibit their usual expressive function when applied to these public corporations.

1 The Intentional Stance

1.1 Introduction

In various accounts on group agency, the underlying assumption is that certain groups are agents because we can interpret and explain their 'behaviour' via the intentional stance.¹ The intentional stance is an explanatory strategy that was introduced by Dennett (1981, 1989). It may be effectively applied to groups because, much like animals and robots, they 'fit the functional profile of agents'.² One of the main questions which this dissertation aims to explore is the extent to which the functional profile of groups actually agrees with the profiles of humans and, especially, of moral agents. This demands looking more closely at the underlying ontology and theory behind the intentional stance. Of particular interest will be the limits and methods which can be utilized in applying it. Although List and Pettit build on the intentional stance, their application of it is quite unlike the way in which it is applied to predict and explain the behaviour of animals and people. Indeed, the two challenges which I will raise in Chapters 3 and 4 can be traced back to List and Pettit's restriction of their methodology in applying the intentional stance.

This chapter will be structured as follows. First, I will summarize the general idea behind the intentional stance (Section 1.3), looking closely at the rationality assumption which is a core part of the tool box. In Section 1.4, I consider some of the intricacies of applying the intentional stance and what means we can avail ourselves of in attributing mental states to a system. Then, I look at the available options when the intentional stance fails (Section 1.5). Finally, I take a step back and explore the ontological status of agency ascriptions within the intentional stance, before summing up the important takeaway of the intentional stance for theories of group agency (Section 1.7).

¹This is true of a number of authors, among them Clark (1983), Tollefsen (2002), List (2012), Pettit (2017), Collins (2019) and Björnsson & Hess (2017). French (1979), who thinks of corporations as Davidsonian agents, may also be counted as a proponent of this strategy.

²See (Pettit 2017, 18).

1.2 Explanatory Strategies

There are many ways to interpret or explain the behaviour of a system or object. In natural science, when we want to predict and explain what happens when we kick a ball, we determine the ball's physical properties, perhaps even to the microscopic level, including its surface friction, its air pressure, and so on, as well as the physical properties of its environment, such as air resistance and wind speed. Having gathered all these facts, we 'feed' them into the laws of physics to predict the outcome – the trajectory of the ball. Dennett (1981) calls this explanatory strategy the *physical stance*. It is perfectly viable in many simple or controlled contexts, but easily gets out of hand in more complex situations. Suppose we wanted to predict the outcome of pressing the 'call' button on a smartphone. Applying the physical stance would demand determining the flow of electrons through various resistors, enormously complex systems of switches directed by sophisticated software, which we would also have to translate into 'hardware facts', electromagnetic waves emitted into the surrounding environment, and so on. It is immediately clear that this process would, in the vast majority of cases, be a completely futile endeavour. Instead, we typically employ a much simpler strategy. We assume that the system, in this case, the smart phone, can be broken up into parts, each of which has a certain design³ and a certain function, and each of which behaves as it is designed to behave. We may understand functions as "capacities that agents assign to parts of entities relative to values or purposes these agents impose on these entities"(typically survival and reproduction for biological entities, and typically practical uses for artefacts).⁴ So, if we see the 'call' button on the smartphone, we will predict that, under normal circumstances, its function is to place a call to the number associated with it. In this case, such a *design stance* is more efficient than the physical stance. However, it is not a viable option if the optimality condition is not met, that is, if the objects under scrutiny diverge from how they usually are. For example, if my phone had just fallen into a puddle, or had been captured and locked by hackers, the design stance will not work: we must revert to the physical stance. But, sometimes, when a system is not designed, or when it has no obvious designer, we must adopt yet another stance: the intentional stance. The intentional stance is a sub-species of the design stance. In it, we do not assume that an object functions optimally, given its design. Instead

³See (Vermaas et al. 2013, 1133).

⁴See (Searle et al. 1995, 13-23) and (Vermaas et al. 2013, 1143), who argue that there are two distinct ways to understand the design stance. The first references the 'designer' and their intended function, be it evolutionary selection or an intentional agent. The second references only the agent that is applying the design stance to a system.

we assume only that it functions as a rational agent.

1.3 Basic Elements of the Intentional Stance

The intentional stance involves two steps. First, one treats the system as a rational agent. Second, one determines what kind of mental states the system ought to have, ‘given its place in the world and purpose’.⁵

Of course, this is not very illuminating unless we specify what we mean by ‘rational’, and what kind of ‘ought’ we allude to when we try to attribute mental states to our system. According to Dennett’s functional theory, beliefs track the truth and desires motivate us to pursue the good. Given this general characterization of those two mental states, which specific beliefs we attribute to an intentional system depends on what that system *ought to believe* or has reason to believe “given its perceptual capacities, its epistemic needs and its biography.”⁶ Similarly, when we attribute desires to a system, we attribute to it those desires which it *ought* to have or which are good for it, given its needs to survive and to procreate, along with the instrumental means to achieve these goals. For animals and humans, this typically means that we desire food, shelter, sex, wealth, power and, of course, many other objects and ends somehow connected to these basic needs. This doesn’t really restrict what desires we might attribute to a system. But whenever we claim that an agent has desires which are not typical for the environmental niche it inhabits, we must give a special story explaining why the agent has these desires. Similarly, whenever we attribute a false belief to someone, we must explain how their error resulted from facts about the environment and the agent’s limited sensory and cognitive capacities.

Thus, we arrive at a picture of agency on which entities like AI’s, corporations, animals and even thermostats can have a mind and act, so long as they are the appropriate targets for the intentional stance. Of course, these entities differ in the kind of rational requirements they are able to meet and how robustly they can meet them across a range of different situations. An entity or system becomes a ‘true believer’ – a true agent within this framework – if we may apply the intentional stance towards it across a wide spectrum of different situations, where these mental states are connected in a rich way to the system’s other mental states and its environment. By ‘rich’, in this context, I mean that the connections satisfy demanding requirements concerning how the system’s intentional states cohere across time, together with facts regarding both how its conative (like desires, goals and intentions) and

⁵See (Dennett 1989, 17).

⁶See (Dennett 1989, 49).

its cognitive attitudes (like beliefs and knowledge) relate to each other at a time and to the world.

Certain theorist have found it irritating that, on this view, our mind is something like a thermostat. But if we consider it as a purely explanatory strategy which captures and orders patterns that we find not just in human agents, but also in many other entities such that these patterns would otherwise escape us, then we have no reason to fear such liberal use of intentional idioms. For the sake of those who remain worried, I will explore some of the challenges for functionalism and possible responses at the end of this chapter. For now, let us turn to take a closer look at the rationality assumption, which serves as an essential step in using the intentional stance.

1.3.1 Rationality

Determining what kinds of desires and beliefs a system has is but a partial predictor of that system's behaviour. One must further assume that the system is rational: by that assumption, we will be able to establish which beliefs are rationally implied by the presently existing beliefs, and what kind of actions would be rational given the entire set of beliefs and desires attributed to the system. Thus, the rationality assumption is not only essential in predicting the behaviour of the system given its mental states. It also plays a role in determining what other cognitive and conative attitudes we should attribute to the system. Rationality explains why beliefs and desires function as they do. It does this by encoding what is constitutive of these states in such a way that is more specific and more substantive than the highly idealized assumption that beliefs and desires, respectively, track the truth and the good. Depending on how one conceives of and spells out rationality, the shortcomings of the agential system, including its limited cognitive and sensual resources, can be taken into account.⁷

So, what is it for a system to be rational? Dennett (1989, 97) offers a very broad characterization, taking 'rational' to be a "general-purpose term of cognitive approval [and associated, conditional] methods of getting ahead, cognitively, in the world". On this pre-theoretical level of rationality, he says, we rely on shared intuitions about 'what makes sense' to determine what is rational.⁸ The idea that rationality may be understood by way of what it is

⁷See Aumann (1997), Gigerenzer & Goldstein (1996), Kahneman (2011), and Kahneman & Tversky (2013) on influential contribution to the vast field of bounded rationality.

⁸See Dennett (1989, 98). This concept of rationality as making sense is close to the core of the intuition which drives Velleman's constitutive account of the aim of agency, which he identifies as the aim of self-intelligibility. But the same concept also comes up in many other theories of rationality under the heading of coherence and rationalization, as

intelligible for an agent to do given their history and place in the world will also serve as the guiding force behind my further development of List and Pettit's theory of group agency into a constitutive theory.

Since Dennett's work on the intentional stance, much has happened in the theory of rationality and reasons, which has since burgeoned into a huge literature. Two main strands have emerged. They may be roughly distinguished as follows. The myth theorists – among them Kolodny (2007, 2008), Kiesewetter (2017), Lord (2018), Poprawe (2020) and Raz (2005) – maintain that rationality consists in responding correctly to reasons. On the other side of the debate, people have argued that rationality can be spelled out in terms of rational requirements, like a requirement for instrumental rationality or a requirement for consistency among one's mental states. This is known as structural rationality. This position is defended, among others, by Broome (2013), Bratman (2018), Ferrero (2014) and Horowitz (2014). There are further distinctions to be drawn, but this much should suffice for our intents and purposes.

Although the intentional stance is compatible with both understandings of rationality, I will follow Philip Pettit, one of the co-authors of the work on group agency from which I will draw most heavily in working out the constitutive theory of group agency. Pettit belongs to the structuralist camp. I, too, will adhere to a structuralist conception of rationality. This means that I understand rationality as issuing rational requirements that, when fulfilled, specify patterns of mental states which are coherent and intelligible and which, when violated, tell us why a combination of mental states is unintelligible. Encoding the concept of rational coherence in this way allows us to capture the characteristic functional roles of different mental states, such as intentions' role as settling practical issues for us or belief's role of tracking the truth. But even within the debate on structural rationality, there is a lot to disagree about. One might believe that different mental states above and beyond beliefs and desires may also play a role. Bratman (1987) has famously argued that intentions are mental states that are not reducible to combinations of desires and beliefs. His reasoning is that combinations of beliefs and desires lack the kind of diachronic properties that are essential to the understanding of our plans and diachronically extended actions. Bratman believes that intentions are non-reducible plan-like states that help us to settle and resolve disputes and extend our agency into the future. Similarly, on the level of groups, mental states such as identification with a collective have been argued to play a significant role in explaining the joint or shared actions of a group of individuals. Without going into detail about these

in MacIntyre (2013).

theories, it is important to note that different rational theories postulate different kinds of entities or concepts.⁹

1.4 Interpreting an Agent

As I've mentioned, Dennett takes beliefs as aiming at truth. This means that, in general, we come to believe those things which are true, given the capacities of our sensory organs and what our goals and aims make relevant to us. Our desires aim at our personal good, given the environmental niche we inhabit. Furthermore, from a first-personal point of view, we arrive at new beliefs by reasoning from existing beliefs via principles of deduction like *modus ponens*, *modus tollens*, as well as principles of abduction and induction. Given that we are rational, reasoning will also allow us to respond to inconsistencies among our beliefs: once we discover inconsistent beliefs, we drop one (or more) of them. The use of this kind of reasoning can, in turn, justify our belief that a person is rational.

From a third-personal point of view, we can also make predictions on the basis of our knowledge that, when an agent believes that 'p' and believes that 'if p, then q' and cares about 'q', she will come to believe that 'q'.¹⁰ Knowing that she believes that 'q', in turn, enables predictions which require her belief that 'q'. Suppose that 'p' stands for 'it is going to rain' and 'q' is 'I get wet'. Coming to the belief that 'q' leads her, together with her desire or end 'to not get wet', to the practical conclusion that she should take an umbrella: that is a necessary means implied by her end.

Reasoning also features in an interesting method of gathering anecdotal evidence which Dennett discusses. The following passage is from a chapter about determining the capacities for higher-order intentional attitudes in vervet monkeys, but the method is much broader in application – it might be used for any kind of investigation of the properties of an intentional system.¹¹

The accretion of unrepeatable detail can be abetted by using the intentional stance to provoke one-shot circumstances that will be particularly telling. The intentional stance is in effect an engine for generating or designing anecdotal circumstances – ruses, traps, and other intentionalistic litmus tests – and predicting their outcomes. [...] Sherlock Holmes was a master of more intricate intentional experiments, so I shall call this the Sherlock Holmes method.

⁹Dennett (1989) calls these abstracta.

¹⁰See Broome (2013) and his *modus ponens* requirement on beliefs.

¹¹See Dennett (1989, 251) and Dennett (1989, 254, footnote 8).

[...]

It is a particular gift of the playwright to devise circumstances in which behavior – verbal and otherwise – speaks loudly and clearly about the intentional profiles of the characters.

Thus, the core takeaway from the capacity to reason is two-fold. First, reasoning allows an agent to maintain rationality. Second, it allows people to use the Sherlock Holmes method from a third-person perspective to make predictions about what follows from the agent's mental states and whether the agent actually exhibits rationality, together with the set of mental states that we attribute to it.

The importance and role of predictions which we may make thanks to the intentional stance is also made clear in the following passage.¹²

[...] the choice of which attitude-proposition pair to ascribe must be driven, under our functionally oriented approach, only by how the agent behaves, not by its physical make-up. That make-up – be it neural, electronic, or perhaps of another kind – may provide indirect evidence about what performance to expect in different situations, but the performance itself should dictate the representations and motivations we ascribe to the agent.

Two things are noteworthy in this short paragraph. First is the importance of a performance-driven way to apply the intentional stance. Second, and connected to this, is that the make-up of the system plays only an indirect role in determining which mental states to assign to an agent. As we will see in Chapter 2, on List and Pettit's account of group agency, a central pillar of their theory of group agency is a very insightful discussion of the importance of group agents' decision procedures. From there, they arrive at the conclusion that these mechanisms for making decisions, together with the individual group members' inputs into them, alone determine the mental states of a corporation. In doing this, the authors seem to forget their earlier remark quoted above, where they maintain that the substance which makes up an agential system can only provide us with indirect evidence regarding the mental states which we may ascribe to agential systems such as

¹²See List & Pettit (2011, 28). It is indeed the accepted approach that, in applying the intentional stance, we should orient ourselves on the performance of the agent. As Vaesen & Van Amerongen (2008, 783) say: "the choice to adopt the intentional stance does, similarly, not presuppose a pre-established position about the specific goals and beliefs of the agent. These purposes and functions, or goals and beliefs, are rather derived: a successful application of the relevant stance to the entity advances specific purposes and functions, or goals and beliefs."

a group agent. Therefore, we should be careful to keep this litmus test for mental attributions in mind. Although the analysis of the make-up of the relevant system is important and can be a great indication for which specific mental states to attribute in a particular situation, we should not jump to the conclusion that this is the only way to determine which mental states to attribute to the intentional system.

So, we may attribute mental states to a system such as a group agent via observation, reasoning on their behalf, and knowledge of their environmental niche. In the case of group agents, we may also learn something about their intentional profile by studying their decision mechanisms. This will be the focus of our next chapter, in which we will explore List and Pettit's account in depth. I want to briefly mention one final general method of finding out the intentional states of group agents and other complex intentional systems like us humans. Pettit (2017) has argued that we can think of rational groups as conversable agents – that is, agents with whom we may exchange ideas and which are capable of communicating. If this is the case, then such groups may themselves provide us with details about their beliefs and desires which would be otherwise unavailable to us. Given that many intentional profiles are compatible with one and the same pattern of behaviour, this method, which seems to be available in at least certain cases, is especially important. For the method to work, we must already place a greater demand on them: we must expect them to answer us sincerely and without the intention to manipulate us. This seems to require, of group agents, not merely the ability to act rationally, but also to be a moral agent of sorts. As I will argue in later chapters, it is more difficult to meet the requirements of this status compared to the requirements issued by morality.

All of the methods described rely on the often tacit assumption that certain optimal conditions are satisfied. These conditions go beyond mere assumption of rationality, as I have suggested in the case of the conversable agent. We must also keep in mind that, for very complex systems such as structured groups, the context or the favourable conditions and feasible limits of the agent also play an important role in which mental attributions we are justified in making.¹³ Sometimes, agents can form an intention to act according to their mental states, but simply fail in their attempt to do so. Intending to win at a competitive game is one thing, but, surely, doesn't on its own ensure success. At other times, we find that a certain goal is achievable in general, but, due to circumstantial conditions outside the agent's control, is in fact out of reach. We might also simply fall short of certain rational requirements due to our limited cognitive capacities. Sometimes, it turns

¹³See List & Pettit (2011, 20) for some remarks on this.

out that two intentions are not both realizable simultaneously, or that they are based on false or inconsistent beliefs. Indeed, this seems to be the rule, rather than the exception, for complex systems such as human agents.

1.5 Irrationality, Explanatory Gaps and the Design Stance

When, in some situation, we must attribute false beliefs or ‘abnormal’ desires to an intentional system, we must give a special story which references the circumstances or the system’s perceptual or rational capacities, or identifies some other deviation from the idealizing assumptions that beliefs track the truth, that desires target the good and that our mental states are overall coherent.¹⁴

For example, suppose I impute to myself the false belief that I am free this afternoon at 12 o’clock and plan to have lunch around the corner at the university. This belief is false, because I had arranged, just yesterday, to meet a friend for a game of tennis at that very hour, but have forgotten about having done so. Dennett argues that such a false belief, even though inconsistent with the belief that I ought to be at the tennis court – which I would have had were I not as forgetful – will nevertheless play its usual functional role, leading me to stay at the university and not show up at the tennis court. Thus, I will treat the false belief that I am free during lunch as if it was true: I will do so by considering what else follows from that belief and how it combines with the rest of my desires, goals and intentions. Of course, when my friend calls me up at 11, in expectation of our match on the tennis court, I will be surprised, realizing that I had completely forgotten about the meeting and already planned to get some food at the university. I will be relieved about the fact that he reminded me about our meeting, be angry at myself for having forgotten it and quickly drop the intention to have lunch. This case raises an interesting question. Did I really have the relevant intention to meet him to play tennis? Or, did I just have a quasi-intention, or perhaps no intention at all? Even though I lacked the belief that I would meet my friend at the court in an hour, which led me to make plans based on the belief that I would be free and stay at the university for lunch, once I was reminded of my original plans, I quickly reconsidered those of my intentions which were incompatible with my resolve to play tennis at 12.

Just as different theories of rationality disagree about which mental states should be regarded as primitive, they similarly tell different stories about

¹⁴See Dennett (1989, 49).

how these mental states are connected. Thus, the different theories offer different answers to the question whether I had, at 11, the intention to play tennis. Cognitivists, like Broome (2013), Velleman (2009b) and Harman (1986), claim that intention is closely connected to belief. They will argue that there is no determinate answer to whether I truly had an intention to play tennis at 12 before my friend's phone call. They will insist that an intention to X requires you to believe that you will X, which excludes the possibility of forgetting that you will X. Non-cognitivists like Bratman (1987) allow for the possibility that I can have an intention to meet my friend at the tennis court at 12, although I lacked the belief that I will be at the tennis court at 12. Knowing that I tend to be forgetful, I might have set a timer to remind me in advance of our appointed game so that the intention could still perform its typical role of leading me to intend the necessary means, such as bringing my racket to the court.

This case shows that different theories of rationality can make sense of certain patterns, but not of others. Both cognitivists and non-cognitivists alike will label as irrational an agent who claims to have an intention to X, but who does not adopt what she believes to be the necessary means available to do X. But they will disagree with regard to when that agent is irrational. While cognitivists will argue that the agent is being irrational once she has forgotten that she will X, non-cognitivists will allow us more leeway in interpreting her as rational and in attributing an intention to her. Importantly, both camps will find it difficult to make sense of the agent's behaviour once that agent is irrational in the sense of having incoherent attitudes. For example, both will have problems saying that a supposed intention for which one does not end up adopting the necessary means is indeed an intention, since instrumental rationality is a constitutive standard by which both accounts understand intentions. In these cases, they might resort to different, weaker concepts such as desires so that they may still make sense of someone who has a conative pro-attitude towards something but does not intend the necessary means. But this strategy of adding weaker mental attitudes to the picture will only take one so far. At some point, even this strategy will hit a wall, especially if, for example, in the case of the planned tennis match, I keep insisting that my attitude towards playing tennis was indeed an intention, even though I do not intend any of the means necessary. There will be, in Dennett's words, an "explanatory gap".¹⁵

The above cases reveal a tension of the intentional stance when faced with irrationality. On the one hand, we have no other strategy available than to adopt the intentional stance and to assume, in turn, some degree

¹⁵Davidson (1985) comes to a similar conclusion in his seminal paper on irrationality.

of rationality and idealizing assumptions about the relevant mental states if we want to make sense of the agents' behaviour at all. On the other hand, if the agent is really irrational, there will be a significant gap between our idealizing assumptions regarding beliefs and intentions and the intentional states which we will end up attributing to the intentional system. In these cases, there is no fact of the matter with regard to the question of whether I really intended to play tennis at 12, given that what is constitutive of having an intention to *X* is intending the necessary means implied by that intention. When I am 'irrational' – no interpretation at all will be settled on.¹⁶ But even in these cases, the intentional stance might still be the best explanatory strategy available. Dennett gives the following analysis of someone who is irrational.¹⁷

Jones has just spent three months of hard work building an addition to his house. It looks terrible. Something must be done to resolve the uncomfortable cognitive dissonance. Count on Jones to slide into some belief that will save the situation. But which one? He might come to believe that the point of the project, really, was to learn all about carpentry by the relatively inexpensive expedient of building a cheap addition. Or he might come to believe that the bold thrust of the addition is just the touch that distinguishes his otherwise hackneyed if 'tasteful' house from the run of the neighbourhood houses. [...] But which of these is actually believed will be determined by seeing what he says and does, and then asking: what beliefs and desires would make those acts rational? And whatever delusion is embraced, it must be and will be – carefully surrounded by plausible supporting material, generated on the counterfactual assumption that the delusion is an entirely rationally held belief. Given what we already know about Jones, we might be able to predict which comforting delusion would be most attractive and efficient for him – that is, which would most easily cohere with the rest of the fabric of his beliefs. Even in the case of cognitive dissonance, where the beliefs we attribute are not optimal by anyone's lights, the test of rational coherence is the preponderant measure of our attributions.

Understanding cognitive dissonance from within the intentional stance vividly shows that we cannot and should not drop our idealizing rationality assumptions in the face of irrationality. When is it appropriate to drop the intentional stance? We have already discussed that predictability provides us with

¹⁶Dennett (1989, 88)

¹⁷See Dennett (1989, 100).

some general standard of when to employ the intentional stance. We do not adopt it if doing so does not afford us a better way to predict the behaviour of a system, such as the trajectory of a stone I have hurled at a window, than the physical stance. We also do not adopt it if the system, although complex, fails in a systematic manner to conform to the rationality condition. But how to navigate this tension?

First, we might have evidence that it is important for the system to be rational, although it currently fails to live up to the standards of rationality. Let us return to me forgetting my tennis session. Upon seeing my friend all geared up and on his way to the court, I immediately remember my actual plan for the afternoon. I blush and exhibit surprise. This reaction can only be explained by my assumption that I was rational to have lunch at the university, and the fact that this assumption has now turned out to be false. Realizing that my beliefs don't match and caring about my rationality, I feel shame and anger at not being able to live up to my own expectations. Thus, my surprise can function as an indication that I mind whether I am rational, that I am motivated to restore mental coherence. Both of these are indirect evidence that the intentional stance is appropriate, even though it does not yield true predictions.¹⁸ Indeed, this criterion is what List and Pettit take to be the defining requirement for establishing when the intentional stance is appropriately directed at a group. They write:¹⁹

We regard a group as an agent just when we think something is amiss if those attitudes are inconsistent, or otherwise irrational. We assume that only group agents as opposed to mere groups should acknowledge that this is a fault that should be rectified.

Second, there is a different kind of evidence that an agent is rational. In some cases of observed irrationality, we can nevertheless identify features in the agent's environment which 'select' for rational agents. A robot, for example, can be assumed to satisfy the assumption of rationality by virtue of the fact that it has been selected and designed by those who programmed it. On the other hand, an animal that lives in harsh conditions can be assumed to be rational in general, since it would not survive otherwise. When we attribute rationality to humans, we similarly have to point out processes of natural selection that select for rationality. In the case of groups such as corporations, the selective environment which plays the analogous role is a competitive market or the legal framework that structures and shapes the group's agency and which has been set up by an intelligent designer itself.

¹⁸See Dennett (1989, 84).

¹⁹See List & Pettit (2011, 39)

1.5.1 Adopting the Design Stance

Yet, if irrationalities mount and the attributions and idealizing assumptions involved in attribution of rationality ramify and lead to false predictions, the theoriser knows that something important has been left out. What must be added to the idealized story so that the actions are rational after all? One reaction, similar to the strategy above though with a slight twist, is to check how the agent ought to be designed so that, in acting as it does, it counts as rational. This is slightly different from the above strategy, since this strategy puts more emphasis on the kind of theory of rationality that is in the background. It is an explanatory strategy that Dennett advises for the case of monkeys or other systems for which we do not yet know what theory of rationality and what mental states we should ascribe to them. It asks us, in identifying what kind of mental states we should attribute to a system, to identify the purpose, the aim, or the need of the system together with constraining conditions regarding its epistemic capacities and then to design or to reconstruct the system as an engineer would.²⁰ Thus, it means that we adopt the *design stance*.²¹

This role of the design stance is also identified by List and Pettit.²²

In the design stance, we identify aspects of a system's make-up that limit its proper functioning as an agent – in the individual case, *idees fixes*, blindspots, obsessions, and the like. A theory that enables us to recognize when there are group agents and when not also allows us to adopt a design stance towards the group agents we recognize. The history and organization of a group agent often leave it susceptible to certain forms of malfunction and underperformance.

List and Pettit use the design stance to identify when the system's inner structure does not function well or when the environment prohibits it from meeting the requirements of rationality. They use the design stance as a tool that can show us under which internal design or within which environment

²⁰See Dennett (1989, 264).

²¹This strategy is reminiscent of the strategy adopted by Bratman (2000) called *creature construction*. This method draws on the work of Grice (2013, 1975). Indeed, it is similar to Dennett's use of the design stance for biological systems. As Vermaas et al. (2013, 1136) note, 'Dennett seems to include his intentional stance more explicitly into the design stance when he explains the interpretation of biological entities as if they were technical artefacts. Dennett does this by means of the engineering technique of reverse engineering, and describes reverse engineering of a technical artefact in turn as a reconstruction of the reasons the original engineers had when designing the artefact.'

²²List & Pettit (2011, 13)

the system is more likely to meet the requirements of rationality. This use of the stance assumes of a system that its function is to aim for rationality. This follows Millikan's understanding of the design stance as underlying and grounding the intentional stance.²³

[I]f some actual organism had acquired fully rational dispositions to behavior totally by accident, and if the observed appearances of rationality resulting from these dispositions were known by us to have resulted from accident, we would be foolish indeed to project these appearances into the future.

[...]

from design for rationality, one can infer real dispositions to rationality patterns, as opposed to mere temporary illusions of such dispositions. It thus appears that the intentional stance must be underwritten by the design stance, rather than vice versa.

But the design stance can also be used to identify what mental attitudes to attribute to a system. The design stance is the explanatory strategy which we use when we attribute a function to a system and assume that it is optimally built to fulfil this function. As we have now seen, the kind of function which we attribute to a system can be quite abstract – like rationality – or more concrete – like a specific purpose or end such as a knife's function of cutting. When we use it like this, we treat the system more like an *instrument* or a *tool*. I will follow Vermaas et al. (2013) in their definition of the two terms.

If we successfully and efficiently predict the behaviour of an entity with the intentional designer stance, then this entity is understood as an entity designed intentionally by agents. And if we adopt the teleological design stance for an entity, it is understood as an entity that has a functional structure through natural selection or simply by its complexity. Let us now call entities for which we adopt the teleological design stance *instruments*, and entities for which we adopt the intentional designer stance *tools*.

Later, I will argue that we can view certain corporations as tools and instruments, since they are designed, both internally and by being placed in a particular social and legal environment, to fulfil the function of profit-maximizing. Their internal design, due to actors such as the legislation, makes them into tools, while the competitive market environment and other non agential facts can make them into instruments.

²³See Millikan (2000, 61, 61-62). This quote is from Vermaas et al. (2013, 1138).

1.6 Ontology and Critical Remarks

Let us now consider the ontological status of the mental attitudes which we attribute to a system such as a group. In defending the intentional stance, Dennett argues for an instrumental realism. The idea is that we need the intentional stance because, without it, we wouldn't recognize the existence of complex patterns. This is especially true in the case of highly complex neurological systems such as ourselves or socially complex entities like modern corporations. On the other hand, Dennett does admit that mental states such as beliefs do not carve nature at its joints. Nevertheless, beliefs, desires and intentions are real. They exist much like *centers of gravity*, parallelograms of forces and the Equator do. Goals and centers of gravity are 'abstracta rather than part of the furniture of the physical world' and in being attributed in statements that are true only if we exempt them from a certain familiar standard of literality. [...] People really do have beliefs and desires, on my version of folk psychology, just as they really have centers of gravity.'²⁴ Dennett is a realist about mental states, because he thinks that statements about these states may be either true or false, as they point to real phenomena and objective patterns.²⁵ But he also maintains that beliefs and the like are abstracta which do not carve nature at its joints and do not belong to the 'furniture of the physical world'.

Another way to understand the instrumental aspect in Dennett's ontology is to contrast it with the claim made by functional realists. Functional realists, such as Fodor (1975), maintain that mental states have analogous neurological structures. Were we able to identify them, we could simply read the content of our beliefs off of them. Dennett's claim is much weaker: he does not posit unique neurological states associated with a specific belief or desire, but claims that many different processes can give rise to the functionally same mental state.

Functionalism about group agency is the prevailing way to understand the behaviour and actions of groups such as governments, corporations, clubs, churches and NGOs. But it is not the only game in town with regard to the philosophy of mind. Searle et al. (1983), Searle (1992) and Chalmers (1996) are just some of functionalism's many detractors, who argue that it is not the right theory of mind because it gets consciousness wrong. These theorists argue that only entities which are capable of having qualitative experiences can be said to have a mind. These are the intuitions which Ludwig (2017a) and

²⁴See Dennett (1989, 72).

²⁵See Dennett (1989, 37 ff)

Rönnegård (2015) appeal to in denying that corporations can have a mind.²⁶ Indeed, some of the intuition-pumps appealed to in the literature, like Block (1980)'s Chinese state thought experiment, take mechanically acting groups which instantiate patterns that have the functional property of mental states to be a counterexample to functionalism. I am not convinced that a thought experiment alone can shoulder the burden of plausibility for a whole theory of mind. Furthermore, all of these theorists have enormous problems in making sense of our talk of groups, which describes them with intentional idioms. They can't explain why intentional descriptions work so well, nor do they supply us with a workable alternative theory that can avoid understanding groups as independent functional structures amounting to agents.²⁷ Until they have solved these issues, we shouldn't worry about their short-handed objections and may continue to make progress on providing a canonical way to understand group agency.

1.7 The Intentional Stance and Group Agency

The discussions of the workings of the intentional stance will help us to contrast List and Pettit's methodological strategies with how the intentional stance is used in other disciplines. This will reveal some peculiar choices made by them in constructing their account of group agency. Now, let me

²⁶This leads to the interesting observation by Silver (2021), who alleges that these authors tend to dismiss functionalism in a footnote without providing any substantive argument.

²⁷Ludwig (2017*a*) tries to understand corporations as extremely complex webs of proxy relationships and roles of individuals. In the final chapter on business ethics, I argue that this approach strongly misrepresents the functional and normative characteristics of the legal framework in which corporations are embedded and the underlying moral relationships within a public corporation. Silver (2021) tries to give an account of group actions via an account of group reasons which is not built on an account of group agency. This is a peculiar strategy, since he does accept functionalism about reasons. Silver (2021, footnote 37) says that it is non-mental states that instantiate motivating reasons in groups, although given that those states play the same role as reasons, why shouldn't they then count as mental states. Given that they play the same role, this seems to imply that these groups at least have the desire (in the case of non-cognitivist accounts of agency) to act upon these reasons, or that they judge that they should act on them. Either way, the group must recognize that it should act on them, especially in cases where reasons underdetermine what to do, as when the group is faced with an incommensurable choice. How could a group, without having a state which plays the role of settling a deliberative problem for the group – a state functionally equivalent to an intention – deal with an incommensurable choice? For these reasons, I believe that Silver's account will imply the existence of at least some states that play the role of mental states, such as an intention and a state that recognizes the system's reasons, for example, a group belief.

sum up the key lessons from our discussion of the intentional stance which I want to take over to the next chapter.

An essential aspect of the intentional stance is its predictive power as an explanatory strategy. It serves as a standard for when we should assume the intentional stance in the first place (section 1.4). We generate predictions about a system by reasoning on its behalf, and we so reason on the basis of intentional states that we attribute to it. The intentional states that we attribute to a system such as a group are not only determined by the group's inner workings. They can also be anchored in facts outside the group, and are also determined by the environmental niche that it inhabits. For example, the environmental niche of a corporation is typically a competitive market. Considering the environment supplies us with evidence for our assumption of rationality, as well as with evidence for certain kinds of goals and desires for survival (section 1.3 and 1.4). To do this, we assume the design stance to determine whether the designer of a group has created an organisational structure which ensures rationality, or whether the environment, much like nature, selects for rational agents. Similarly, the design stance also works if we want to find out whether the group has been designed towards a specific, substantive goal, like a corporation designed for making profits or a sports club making it possible for people to play football (section 1.5 and 1.5.1).

The assumption of rationality is also warranted given other kinds of evidence for rational agency. The presence of the capacity to reason typically indicates the system's rationality to be more robust than it is for systems where that capacity is absent. Another type of evidence for rationality appears if the system exhibits 'surprise' and other affective attitudes when it does not live up to basic standards of rationality (section 1.5). Although it is an essential part of the intentional stance that we assume rationality, irrationally held beliefs do not force us to drop the intentional stance. This is another important issue not addressed in List and Pettit's rationalized account of group agency.²⁸

²⁸It is included in Björnsson & Hess (2017)'s theory of group agency.

2 A Thin and Thick Account of Group Agency

2.1 Introduction

In this chapter, I will explore one of the most prominent accounts of group agency in the debate developed in List and Pettit's (henceforth L&P) 'Group Agency – The possibility, design, and status of corporate agents'. Since they build their account of agency on the intentional stance, the two main questions which will occupy us for the first part of the chapter concern how to attribute mental states to groups and under which conditions these groups satisfy rational requirements. In addressing these questions, I will lay out some of the ground work and tools that I will later draw upon for my own extension of their account, a constitutive account of group agency. One of the guiding questions of my dissertation is whether groups can 'perform effectively in the space of obligations', which can only be answered given an account of a group will.¹ L&P understand this question as asking whether or not groups can be persons, but we may also understand it to be a question about the moral agency of groups, contrasted with the mere agency possessed by young children and animals.² As a result, they provide a thin and a thick account of group personhood. The thin account requires of groups that they can form intentional attitudes about normative propositions and act on them. The thick account demands, further, that they are capable of reasoning, that they have a rich capacity for self-regulation and that they see themselves as having that very capacity. Christian List believes that, to be a competent party in the space of obligations, or to participate in moral relationships and be capable of good or ill will, group agents only need to satisfy the thin account, while Philip Pettit believes that they have to satisfy

¹See List & Pettit (2011, 173, 177) and Chapter 4 on why a group will is essential for groups to be part of a moral community shaped by obligations and entitlements stemming from reciprocal obligations.

²Personhood goes further than moral agency, since it also includes the rights that one has against others and their moral status, that is, questions about dignity and intrinsic value. For our purposes, it is enough to consider the question about personhood to be one of moral agency and the capacity to have a good or ill will.

the thick account. In the literature on group agency, this difference between the authors is often overlooked and their work is typically associated only with the thin account, forgetting the fact that one of them offers a more demanding alternative. I think that it is helpful to look at both accounts, especially since one of them captures the idea of a good or ill will by a group agent much better than the other.

The chapter is organised as follows. I start out by distinguishing structured groups, the main target of this investigation, from groups that do not rise to the level of agency (Section 2.2). Then, I analyse how a group's hierarchy or organisational structure can give rise to or prohibit the satisfaction of rational requirements and how these organisational structures can help us identify what mental states we may attribute to a group agent (Section 2.3). A key concept in this discussion is the notion of an aggregation function or mechanism. As this key concept has received only little attention in the literature so far, even though it plays an important role in some of their methodological and ontological claims, I take a closer look at what underlies it and how this can be problematic (Section 2.4.2). Then, I discuss what I take to be the account of action associated with the thin account of group personhood (Section 2.4.4). I then turn to Philip Pettit's thick account, which I explore by first looking at the notion of group identification (Section 2.5.1) and the more demanding notion of group self-identification (Section 2.5.2). In Section 2.5.3 I consider whether, under the thick account of personhood, a group can be said to be capable of a will. I show that Pettit's account lacks the reasoning-centered aspect of an account of a will, which is why we have to add something to the existing framework that L&P have proposed. What needs to be added, I argue in Section 2.4.3 is an account of group deliberation that is inspired by Michael Bratman's theory of shared deliberation. I conclude the chapter by looking at the metaphysical foundations the slightly modified account and why we should think of these and similar accounts as proceduralist theories of group agency (Section 2.6).

2.2 Collections and Group Agents

What is it for a group to be an agent? In our previous chapter, the answer to this question amounted to adopting the intentional stance towards it, in such a way as to robustly provide us with accurate predictions. The intentional stance, as an explanatory strategy, requires us to attribute mental states such as beliefs, desires, goals and intentions to a group. The group must be able to form representational states that depict how things are in the environment that it inhabits. It must have motivational states that specify how the agent

requires things to be in the environment and it must possess the capacity to process both of these kinds of states in order to appropriately intervene in the environment, whenever that environment does not fit the motivating specifications.³ Once we have attributed mental states to a group, our theory of rationality tells us what the group will do. For example, knowing what a group's ends are and what it takes to be the necessary means to these ends allows us to infer the course of action the group will choose.

Suppose we were to impute some desires or beliefs to *collections* such as ethnic groups, people who share a certain nationality, millennials, white- or blue-collar workers, or people who are sitting in a tram. More specifically, take, for example, the collection of people waiting at a particular bus stop at a particular moment. Assume, further, that we attribute to them the desire to get onto the bus as a group, along with the belief that, by getting onto the next bus, they will realise this desire. This intentional explanation is not robust: it does not provide us with reliable predictions under only slight counterfactual deviations. It is easy to imagine that someone among the crowd at the bus stop is only there because she wants to pick up a friend, while someone else is just studying the timetable. Similarly, if, for lack of space, some of those waiting at the bus stop couldn't get onto the bus, this would not lead everyone else to step out in solidarity with the hope that the next one would be able to fit all of them in. Of course, we could easily come up with another, more specific desire – perhaps the group intends to get into the next *two* busses if the first one cannot accommodate the entire group. But this desire would not cover the intentions of our bus-plan reader or the caring friend. The success of the intentional explanation would be *ad hoc* and, more importantly, even if true, would not involve any other desires and beliefs which go beyond entering the bus. It wouldn't be a good explanation for this kind of social phenomenon. Hence, we need a more systematic way to identify the right kind of groups for which the intentional stance represents a good explanatory strategy, as well as a more principled way to attribute mental states to a group agent.

2.3 Aggregation Functions

One such way has been explored in depth by List & Pettit (2011), who defend an account of group agency in which the mental states that we attribute to a group can be identified by considering aggregation functions or, more intuitively, *decision procedures* which aggregate the individual mem-

³See List & Pettit (2011, 20–21).

bers' intentional attitudes and output the group's attitudes. They do this by building on the results developed in social choice theory on the aggregation of attitudes. This way of understanding collective mental states has certain ontological and methodological consequences that we will consider more closely in Section 2.6. The benefits of this approach are clear. By relying on the formal tools and frameworks of social choice theory, we can express certain results more precisely and transparently. Furthermore, we can easily transfer the results already generated in said field and apply them to our account of group agency. On the other hand, what we gain in simplicity and transparency, we lose when we try to apply these concepts to the social reality of corporate agents. For one, finding the social entity corresponding to the mathematical notion of an aggregation function turns out to be far from straightforward. Are their decision mechanisms embedded in an organisational hierarchy or structure? Or can we apply the notion of an aggregation function in a more permissible way to any kind of individual contributions that give rise to patterns on the level of groups that can be explained by the intentional stance? I will consider these questions in Section 2.4.2.

Nevertheless, L&P's use of aggregation functions has proven highly influential and a great way of showing some very interesting results for group agents. Among these is showing how group mental states are independent from their members and what kind of properties a decision procedure must have to make it more likely that a group agent will satisfy rational requirements.

2.4 The Thin Account of Group Agency

To give an account of how the individual attitudes of the members of a group agent relate to the group agent's collective attitudes, L&P first narrow down the kind of individual attitudes on the basis of which a group might form its attitudes. These are binary beliefs or preferences towards propositions. They are binary in that they can be either positive or negative. So, in a group of two people consisting of Ann and Bob, if Ann believes the proposition p = 'that the kitchen is dirty', then, in List and Pettit's language, she can be said to have a positive attitude towards the proposition p . If Bob believes that the kitchen is not dirty, then we say that he has a negative attitude towards p . Preferences are treated similarly. Preferring that p be the case, rather than not, is understood as a positive attitude towards p , while an agent who prefers that not p be the case has a negative attitude towards p . Apart from atomic propositions like p , individual agents can also believe any formula that can be formed by combining propositions via the Boolean connectives

\wedge and \neg . The combination of the attitudes held towards a given formula by all the members of a group is called a profile.

An *aggregation function* takes as input the profiles of individual members' positive or negative beliefs or preferences with regard to a formula φ and aggregates them, for example through a majority vote or a dictator's decree, to form the corresponding positive or negative collective group attitude towards φ as an output.⁴ Such an aggregation function can be enormously complex, involving a transformation of individual attitudes, the assignment of different weights to the attitudes of different individuals and a kind of collective reasoning which guides the transmission from one set of collective attitudes to another. But it can also be a very simply mechanism. Suppose, for example, that a corporation's board of managers considers whether to expand their business to China. To do this, they will put the proposition 'that the company will open a factory in China' on the *agenda* – a list of several propositions on which the company must take a positive or negative stance – and come to a collective decision via a majority vote.

One of L&P's main contributions to the literature was to point out that not every aggregation mechanism that satisfies plausible initial assumptions about propositional inputs and their connections can guarantee that the resulting group attitudes will form a consistent and coherent whole and allow us to adopt the intentional stance towards the group agent. This is so even if we assume that each of the individual members themselves holds rational or consistent attitudes.

L&P's example takes an initially very plausible aggregation mechanism – majority voting, in which a group adopts all those attitudes that are held by the majority of its members – and shows that in the case of simple, logically interconnected formulas, the group's attitudes turn out to be inconsistent even if each member is individually rational.

This example, known as the 'discursive dilemma', can be nicely illustrated as follows. Suppose that the three-member board of directors of a company in the energy industry must decide on whether to invest into fracking. Fracking is a technology that allows oil and gas to be extracted from layers of underground sediments by pumping a mixture of harmful chemicals into the ground. Although this technology has proven environmentally harmful, increasing the chance of earthquakes and poisoning ground water, it is also very profitable. Suppose, also, that the pay of the company executives

⁴See List & Pettit (2011, Chapter 3) for a number of different aggregation functions such as majority voting, dictatorial voting – in which the judgements of a single dictator determine the group's attitudes – or other more informal mechanisms such as collective deliberation. I will be using the terms 'aggregation functions', 'aggregation mechanisms' and 'group level decision procedures' interchangeably throughout my dissertation.

is linked to the company's profitability. To inform their decision, the three board members must vote on four different propositions regarding what they believe the company ought to do and by what means. They are:

1. The company ought to pursue the strategy which maximises profits.
(P)
2. If the company wants to maximise profits, it must invest into fracking.
($P \rightarrow F$)
3. If the company invests into fracking, the environment will be harmed.
($F \rightarrow Q$)
4. The company ought to invest in fracking and, as a side effect, harm the environment. ($P \wedge Q$)

Suppose that the board members vote as shown in Table 2.1 below. It is easy to see that each individual's beliefs are consistent, while the group's beliefs as a whole are inconsistent. The group endorses each of premises 1-3, from which it logically follows that it believes that it should maximise profits and, via *modus ponens*, believes that this will lead to environmental pollution. So, the group must also believe that it ought to maximise profits and, as a side effect, harm the environment. Nevertheless, since none of the individual members of the board accepts this conclusion, the firm, because it determines its attitudes via majority aggregation, does not accept it either. If we assume that consistency between beliefs is a rational precondition for adopting the intentional stance towards a group, then an aggregation function such as majority voting, which leads to the firm holding inconsistent views, does not robustly ensure that we can interpret the firm as an agent.

The most straightforward way to avoid this is to adopt a different aggregation function. This is L&P's conclusion. They discuss several aggregation mechanisms with regard to the following conditions or characteristics. These

	P	$P \rightarrow F$	$P \rightarrow Q$	$P \wedge Q$
CEO 1	Yes	No	Yes	No
CEO 2	No	Yes	Yes	No
CEO 3	Yes	Yes	No	No
Firm	Yes	Yes	Yes	No

Table 2.1: A discursive dilemma

are the characteristics which they assume an aggregation function, at least *prima facie*, ought to satisfy.⁵

Universal domain. The aggregation function admits as input any possible profile of individual attitudes towards the propositions on the agenda, assuming that individual attitudes are consistent and complete.

Collective rationality The aggregation function produces as output consistent and complete group attitudes towards the propositions on the agenda.

Anonymity All individuals' attitudes are given equal weight in determining the group attitudes. Formally, the aggregation function is invariant under permutations of any given profile of individual attitudes.

Independence The group attitude about each proposition depends only on the individuals' attitudes towards it, not on their attitudes towards other propositions.⁶

Consistency, relative to the agenda, refers to a set of propositions that are co-realizable. Completeness means that the collective agent has, for each formula in the agenda, a positive attitude to either the formula itself or to its negation. The idea of rationality that List and Pettit employ here is a set of rational requirements, such that an agent is irrational if she violates any such rational requirement. Understood this way, collective rationality means that rationality requires that, if an agent believes that X, then she does not believe that not X. Another requirement, utilised above but not mentioned among the four conditions, is a *modus ponens* requirement. It states that rationality requires that, if an agent S believes that p, believes that if p, then q, and cares about the truth of q, then S believes that q.⁷

The first two conditions specify the admissible input and the 'ideal' output. The first condition allows for groups with diverse opinions to be able to reach a shared conclusion and, thus, demands of the aggregation mechanism to be applicable to a wide range of topics that might also prompt conflict. On the other hand, collective rationality is of special importance, in L&P's view, since it ensure that the group can adopt a practical standpoint towards the problems and tasks which it faces, and can act on any such standpoint.

⁵See List & Pettit (2011, 49).

⁶In their book, List & Pettit (2011, 55) actually require a stronger condition called systematicity. But, since systematicity is only a strengthened form of independence and is given up completely for the aggregation mechanisms which they discuss in detail, I will only explore this weaker condition here.

⁷Broome (2013) proposes a similar requirement.

This possibility for action would be unavailable if the group had inconsistent beliefs or intentions, because, in that case, it would not be clear on which beliefs or intentions it should act. Therefore, the second condition must be satisfied if we want to adopt the intentional stance towards the group.

The final two conditions specify how input and output ought to hang together. Anonymity specifies that each individual should have the same say in the decision. This condition plays a significant role in democracies or groups in which we want to utilise as much of the information, spread across the members, as possible. It is violated in groups where certain of the group's beliefs are determined by experts. In the example above, the group might assign the task of determining whether fracking harms the environment to a geologist whom they employ or even to a person external to the group. Importantly, if the aggregation mechanism is implemented by an outside authority, then even though the group attitude will be determined by someone external to the group, whether the group actually has the relevant attitude will still depend on the members' contributions with regard to this attitude.⁸ Groups which are organised hierarchically, like corporations, sport teams led by a coach or states, violate anonymity: the people at the top will either be the only subgroup that determines the group's overall direction or their vote will have a greater weight relative to members from other subgroups.

Independence, on the other hand, is the idea that the only thing relevant to collective attitudes with regard to a proposition φ are the corresponding individual attitudes about φ which are plugged into the aggregation function.

List & Pettit (2002) have shown that there is no aggregation function that satisfies all four of these conditions. The discursive dilemma establishes that a majority function, that satisfies the conditions of universal domain, anonymity and independence, does not satisfy robust rationality because it generates inconsistent attitudes on the part of the group. The authors also discuss several aggregation functions which preserve robust rationality in discursive dilemma cases, but do so at the cost of violating one or more of the other conditions. For my purpose, the most interesting aggregation functions that ensure collective rationality in discursive dilemmas are the functions which give up independence. These functions are of particular interest because they demand that the collective collectivises reasons or treat group attitudes like group reasons. Therefore, these are the functions which are typically at play in rationally robust group agents.

⁸See List & Pettit (2011, 221, endnote 62). In these cases, it is clear that, as in the premise-based mechanisms which will be discussed later, no member of the group must share the attitude that is determined by an outside authority. For the attitude to function like a belief, it needs only to issue in the right kind of action on the group level. For this, the individual enactors need not share this belief.

Sequential priority procedures

Aggregation functions that relax independence prioritise some propositions over others. One such aggregation function is a ‘premise-based procedure’, whereby the group generates its attitudes about the premises by means of a simple majority of votes and derives its conclusions from its attitudes towards those premises. In the discursive dilemma, discussed above, this would mean that the group only votes on three questions: whether it ought to maximise profits, whether the necessary means for this end include fracking and whether fracking entails harming the environment. Since the group affirms all three formulas, the premise-based procedure will entail that the group will also affirm the conclusion entailed by these premises – the company ought to invest into fracking and, thereby, harm the environment. Alternatively, the group could also decide to adopt a ‘conclusion-based procedure’, voting only on the conclusion without forming attitudes about the premises. One problem for such an aggregation function is that it is not complete, since the group holds no attitudes towards the propositions which it labels as premises on its agenda.

These two aggregation functions share an important aspect: both treat one set of attitudes as reasons for or against other attitudes. In the case of the premise-based procedure, the group derives its conclusion from the reasons (pr it accepts. In the case of the conclusion based procedure, the attitudes labeled as conclusions provide the group with reasons not to adopt other attitudes. In this sense, both aggregation mechanisms ‘collectivise reasons’. But the collective reasoning must include more than this, because the group must also decide which propositions should count as premises and which should count as conclusions. Settling this question is not at all trivial – the difference that it makes is nicely illustrated by the saying ‘One man’s *modus ponens* is another man’s *modus tollens*’. Suppose that, in my above example, the CEOs take as a premise the conjunction that they should ‘maximise profits and harm the environment’ ($P \wedge Q$), in addition to the premises that ‘maximizing profits demands investing into fracking’ ($P \rightarrow F$) and that ‘investing in fracking harms the environment’ ($F \rightarrow Q$). The conclusion is the proposition to maximise profits (P). If this were the case, then, given the votes presented in Table 2.1, the premise-based procedure would imply that they should not maximise profits.

But how can one decide which proposition should count as a premise and which as a conclusion? One way to settle this question is by yet again applying an aggregation function such as majority voting. But such a mechanical solution does not seem to be in the spirit of attitude aggregation, where the premises function as collectivised reasons that can be used to justify and

explain a group's standpoint on a certain issue. Furthermore, blindly using a majority vote to determine the roles played by certain propositions will easily lead to inconsistencies over time. These inconsistencies will arise for the reasons outlined above, namely, that it matters which propositions count as premises. Moreover, this purely formal characterization does not allow members to distinguish between premises and conclusions in any explanatory or intelligible way. Should groups count decisions or votes made in the past as premises? Such a view would produce overly rigid mental states for groups that do not reflect how changing circumstances might undermine the assumptions that served as the bases for these votes. Indeed, L&P's collectivised reasons fail to portray reasons' typical roles of rationalizing behaviour, allowing us to describe agents as responsive to reasons and to hold them responsible for acting on some reasons rather than others. As it stands, sequential priority procedures still fall short of allowing us to treat a group as collectivizing reasons that play the characteristic roles of reasons.

A more promising approach, as suggested by L&P, is to aggregate the attitudes of the members by more informal ways, for example by collectively deliberating on the question which propositions ought to count as premises. But can deliberation still be modelled as an aggregation function? List and Pettit argue that it can. Since group reasoning, group reasons and a group acting on group reasons will all become important later on, it is worth spelling out in detail what it means for a group to deliberate and reason, and what concepts are involved in reasoning about and acting on moral propositions. This will also demand looking closely at the notion of an aggregation mechanism. What counts as an aggregation mechanism? What is that mechanism's ontological status? To answer this question, we must investigate a concept that will play an important role throughout the dissertation: shared intention. Once we have done these two things, we will return to the idea of a reasoning group agent and, ultimately, to the key concept of an action by a group agent within L&P's thin account of group agency.

2.4.1 Shared Intentions

Shared intentional activities are an important building block for many parts of this dissertation. They play a role in an account of aggregation functions, in shared deliberation and in my later account of social practices understood via constitutive rules which shape the mind of group agents such as a corporation. Thus, it is important that I am explicit about what I refer to in talking about shared intentions or shared activities, especially since there are competing accounts of the phenomenon of shared activity. In this dissertation, I will be employing the most prominent individualistic account in the debate,

which was developed in a series of papers by Bratman (1992, 1999, 2013). I do not believe that any of my arguments hang on this particular choice: alternative theories by Gilbert (1987, 1990, 1992), Alonso (2009, 2016) Tuomela (1995) or Searle et al. (1983) might support the same claims. Nevertheless, I will stick to Bratman's account, as it is one of the most thoroughly worked out accounts of joint activities.

Bratman develops his theory of shared intentions on top of his account of the planning theory for individual agency. According to his theory of individual agency, intentions are plan-like states that are restrained by synchronic rational norms of consistency and coherence, as well as diachronic norms that govern the stability of intentions over time. These norms put rational pressures on individuals' attitudes and, when functioning properly, allow agents to resolve and settle practical issues and coordinate among different plans.⁹ They also make it possible for an agent to be self-governing over time (or at some time).¹⁰ Beginning from this understanding of individual intentional activity, Bratman develops his account of shared action or, as he calls it, shared intentional activity, by considering the job or role of shared intentions in our lives. He identifies the following set of functions: '[1] It helps coordinate our intentional actions; [2] it helps coordinate our planning; and [3] it can structure relevant bargaining and shared deliberation. And it does these three things in ways that track the goal [of the shared activity].'¹¹ Bratman's conjecture is that shared intentions that play these three roles or perform these jobs are realised by the following set of individual intentions and interrelations between the participants.¹²

Definition 2.4.1. *We intentionally J if*

- (i) *we each intend that we J.*
- (ii) *we each intend the following: that we J by way of the intentions of each of us that we J.*
- (iii) *we each intend the following: that we J by way of sub-plans of each of our intentions in favour of J-ing that mesh with each other.*

⁹See Tenenbaum (2018) for an overview of these functional roles of intentions.

¹⁰For a recent discussion of the issues related to the value of self-governance realised by intentions, see Bratman (2018, Chapter 4, 6-8, 10-11).

¹¹See (Bratman 1999, 112).

¹²See Bratman (2013, 84). Bratman (1993) argues that conditions (i)-(iv) are already sufficient for an account of shared intention. Since some of the uses of the concept of shared activities – like collective deliberation and social practices – extend over long periods of time, the state that realises these phenomena must be sufficiently stable and robust across time to explain those activities' diachronic feature. This is the reason why Bratman appeals to the following set of conditions when explaining shared deliberation.

- (iv) we each believe the following: if each of us continues to intend that we *J*, then we will *J* by way of those intentions.
- (v) we each believe that our intentions in (i) are persistence interdependent.
- (vi) our intentions in (i) are persistence interdependent.
- (vii) there is common knowledge among us of conditions (i)–(vii).
- (viii) public mutual responsiveness in sub-intention and action, mutual responsiveness that tracks the end, intended by each, that we *J* by way of the intentions of each of us that we *J*.

Conditions (i)–(vii) together constitute Bratman’s account of shared intention. He calls condition (viii) the connection condition, which, in combination with the others, realises shared activity. This explanatory relation between the phenomenon of shared intentional activity and the conditions (i)–(viii), together with his account of individual planning agency, the rational norms involved in this as well as appropriate social norms together constitute what Bratman calls the *basic thesis*.

A few words of clarification are in order. As it is used in (i)–(iii), ‘we’ does not already imply a certain form of collectivity. If it did, the account would be circular. As the pronoun appears as the subject of the different conditions, it is meant distributively, referring simply to the people sharing intentions.¹³ The ‘we’ appearing in the content of the intention is similarly meant as a neutral description, not presupposing any collectivity or cooperative character of the intended actions.¹⁴ Imagine, for example, that the members of a group intend that ‘we publish the paper’, which is in the content of each persons’ intention. Then, the content of the intention in condition (i) refers to the outcome of the activity of paper-writing, but does not imply that this is a cooperative activity.¹⁵

¹³The people participating in the shared intention are those to whom the conditions (i)–(viii) apply.

¹⁴Petersson (2007) has argued that ‘we’, as it figures in the content of the intention, is not actually a subject-neutral term and already contains a more substantive notion of we or us: a plural subject.

¹⁵Of course, this is more difficult to imagine in the case of an activity like dancing together, which seems to be, by definition, cooperative in nature. But even in such cases, we might think of the motions that both dancing partners go through as being rehearsed in such a way as to not need a partner. For a deeper discussion of this claim, read Bratman (2013, 64 ff).

The expression ‘by way of’ in conditions (ii) and (iii) fleshes out the idea that we treat each other as co-participants of the shared action.¹⁶ It expresses this relation between us in two ways. First, it ensures that our intentions interlock: that ‘each would continue so to intend if, but only if the other were to continue so to intend’.¹⁷ Put differently, in sharing an intention to φ , each individual must reference the other’s intention to φ , thus relating one’s intention to the other person as a co-participant. Second, in settling a deliberative question by forming a shared intention, I settle the matter not only by myself alone. The matter is also settled by my co-participants having the relevant intention, which supports mine and vice versa.¹⁸

Finally, condition (vi) stipulates that such interlocking of intentions must be persistent until the shared intention is satisfied. This persistence depends on the reasons of each individual for the shared intentional activity. These reasons can take three forms.¹⁹ In cases where one enjoys participating in the shared intentional activity, *desirability-based* interdependence leads to the persistence of the shared intention. In cases where one needs the other to achieve one’s intentions, the interdependence in question is *feasibility-based*. Finally, in cases where one assures, agrees or promises the co-participant that one will stick with the shared intention, or where one has other moral reasons to participate in the shared intention, persistence is due to *obligation-based* interdependence. When all this is out in the open and thus common knowledge (vii), this web of intentions and interrelations will produce a shared activity. Since I will also be applying this framework to ‘massive shared agency’,²⁰ in which such webs span across a large part of society, we might want to weaken (vii) such that it demands only that this web of intentions is publicly accessible.

This is the basic account of shared intentional activity with which I will be working throughout the rest of this dissertation. Before we continue, I will respond to three further issues that might be raised regarding my usage of Bratman’s account of shared intention.

1. How do shared intentions come about?
2. What can be in the content of shared intentions?

¹⁶See Bratman (2013, 50). The notion of co-participation also plays a central motivating role for the introduction of the concept of plural self-awareness developed by Schmid (2009).

¹⁷See Bratman (2013, 49 and 69).

¹⁸This idea – that one may settle some matter for someone else, thereby threatening the own-action condition, which states that one can only settle those things that are up to you yourself – has also been a point of contention.

¹⁹See Bratman (2013, 70 ff).

²⁰See the Bratman-inspired account of Shapiro (2014) for large scale social activity.

3. Is Bratman's account too demanding for most social phenomena?

Let's begin with the first question. The aetiology that can give rise to such a web of shared intentions may take two rough forms.²¹ First, elements or salient features in the common environment might trigger the formation of the relevant intentions, given that the participants share certain assumptions about the world regarding these elements. For example, a song played at a festival might function as a trigger for people to start dancing in a certain way together. Or, as I will argue later on, a certain kind of market environment can bring people to participate in the practice of representing the public corporation as having the goal of making or maximizing profits. Second, in cases in which I know the character of the other person with whom I will form a shared intention, I might correctly judge that she holds a conditional shared intention which would become unconditional once I signal to her that I will form the relevant intention. Or, the other way around, I can predict, knowing her, that if I first signal to her that I have a conditional or unconditional intention of the relevant form, that she will correctly interpret that signal and form the matching intention herself.

What about the possible contents of shared intentions? The typical cases discussed in the literature – walking together, painting a house together or travelling somewhere together – might be taken to suggest that shared intentions can only be used to explain relatively simple and practical activities. But the basic thesis about the nature of shared intention is a much more versatile instrument which may be used to explain much more of our social world. Bratman himself uses it to explain shared deliberation, which he understands as intending that we give a fixed normative weight to some proposition in a decision problem.²² He has also recently used it to understand decision procedures, and his method will be relevant to our later analysis of aggregation functions. Hence, the basic thesis may apply to activities that have a much more abstract content than our walking together.

Such rich applicability might prompt a worry. Bratman's account of shared intentions owes its power in explaining diverse tasks and jobs to its relatively high demands on its participants. Participants must have intentions which are interlocking, reflexive and mutually responsive to one another, while all of this must also be common knowledge. This has led some to conclude that Bratman's concept is too demanding.²³ This argument might become especially pressing as greater numbers of participants become

²¹See Bratman (2013, 73 ff).

²²I will come back to this notion of shared deliberation when I discuss Frank Hindriks' account of a normative perspective.

²³Pacherie (2011)

involved in a shared activity. Bratman has responded that demandingness is difficult to measure empirically and, on the other hand, the whole web of interconnections and intentions need not be actively represented.

The agents need to have relevant intentions that ground associated dispositions of tracking, support, adjustment, and responsiveness in thought and action; and this web of intentions and associated dispositions needs to support the attribution of intentions with the complex content cited in the intention condition. But this need not involve explicit conscious awareness of this complex content. Further, we can also allow that the intentions cited by the basic thesis need not be actual causes of relevant behavior but function only as what Philip Pettit calls ‘standby factors’ that exercise ‘virtual’ rather than ‘active’ control.²⁴

Finally, the demandingness objection is typically leveled against uses of Bratman’s account of shared intentions to analyse cooperative phenomena among children and animals.²⁵ Since we will be dealing with institutions made up of adult humans, these objections are not relevant to his project.

2.4.2 A Closer Look at Aggregation Functions

With the concept of shared intentionality at our disposal, let us return to the question of how to understand aggregation functions. Aggregation mechanisms are the tools by which we assign beliefs, desires and intentions to a group agent. But, as we have already seen, they may become very complex, for example, when, in a premise-based procedure, they take into account decisions made a long time ago as premises, or when they have as input meta-decisions, such as which propositions are to count as premises and which as conclusions. But what is the nature of an aggregation mechanism and how do we identify them?

Shared Policies of Procedure

Typically, we identify aggregation functions by considering the hierarchy of a structured group agent. Hierarchies, what L&P call organisational structures, are the thick and practical variant of the aggregation function: they specify who decides on which issues, how they do it, how these decisions are

²⁴Bratman (2013, 104)

²⁵See Tollefsen (2005) and Pacherie (2011).

constrained by the higher levels of the hierarchy and how they are enacted.²⁶ These hierarchies are normally set up in a shared intentional way by some sub-set of the group agent's members. L&P point out that it is sufficient for this sub-set to participate in the shared intentions for the group to exhibit an overall structure that gives rise to group beliefs, desires and intentions.²⁷ Shared intentions license members to act on behalf of the group and to authorise the group to speak and act on their behalf. The licensing relationship, which fixes a member's role of authorization or activity, plays the role of a membership relation. A group in which only one person is active and decides on behalf of the group, while all the others authorise that person to do so, is a dictatorship. In a democracy, everyone participates in the decision making and everyone authorizes everyone regarding certain decisions. The shared intentions in the background that distinguish both forms of decision making structures are shared policy of procedure.²⁸

Bratman understands shared policies of procedure as a shared policy that favours the operation of a given procedure – such as a majority vote, shared deliberation or the decision of a single person – plus follow-through that is guided by that procedure's output. This must be a shared policy rather than a shared intention, since shared intentions are one-off instances of our acting together, whereas a hierarchy is supposed to provide an organization with a robust, reliable and consistent way of making decisions. Policies, in Bratman's account, are triggered or instantiated if a set of conditions is satisfied. In an organisation, these conditions typically include the type of decision that must be made, as well as other shared policies of procedure which might delegate a certain decision to a sub-group, where the rules by which these groups decide are also governed by a shared policy of procedure.

Most of the aggregation functions which L&P have in mind seem to involve these kind of shared policies of procedure. For example, they say that 'they each intend that they together act so as to form and enact a single system of belief and desire, at least within a clearly defined scope; they each intend to do their own part in a salient plan for ensuring group agency within that scope.'²⁹ Admittedly, this quote is about forming a group agent in gen-

²⁶See List & Pettit (2011, 60). This is basically the idea of a Corporate Internal Decision (CID) described by French (1979, 1984). The CID determines (1) an organizational or responsibility flow chart, which identifies individuals' roles and the interdependent lines of responsibility within the corporation and (2) corporate decision recognition rule(s), for example, what counts as a company's decision (usually embedded in something called corporation policy).

²⁷List & Pettit (2011, 34-36)

²⁸Bratman (2017, 41-43)

²⁹See List & Pettit (2011, 34)

eral. But, given their assertion of the importance of attitude aggregation, where this aggregation refers to voting *procedures*, I take it that, in this paragraph, the authors describe the realisation of a web of sensible aggregation functions that satisfy rational requirements. Similarly, when they describe licensing others to make decisions on behalf of the group and authorising other group members to make decisions that are to count as the group's decisions, they identify the requirement of 'informal acceptance by others', or 'endorsement' by at least some members of the group agent.³⁰

This interpretation can also be found in a recent paper by Pettit. There, he writes:

In decision-theory terms the organization has a kinematics that governs how it would update in such counterfactual situations, with individual members making the psychological adjustments that support appropriate responses. This kinematics is grounded in a central, decision-making structure [...] albeit a structure that may operate via a widely distributed network of sub-structures.³¹

This view is echoed by Collins & Lawford-Smith (2016) and Hindriks (2008). Collins and Lawford-Smith talk straightforwardly about decision-making procedures, rather than the more abstract notion of aggregation procedures. Hindriks, on the other hand, talks about decision procedures which must be *accepted* by the group. All of these versions of aggregation functions, I think, can be fleshed out via the notion of shared policies of procedure.

Further Notes on Aggregation Functions

The intentional interpretation of aggregation functions as webs of shared intentions between group members who specify relevant procedures might be in tension with the fact that L&P understand the inputs for aggregation functions as *manifested attitudes*.³² This property of aggregation functions has also been highlighted by Hess (2012). He argues that, although L&P often talk about attitudes being aggregated, what is being aggregated are the members' manifested attitudes or their actions, which do not necessarily correspond to their genuine beliefs or preferences at all. I agree: this point is not clear enough and, furthermore, it must be abandoned when L&P introduce their thick notion of a corporate person, which relies on the mental concept of group self-identification. We will keep this constraint in mind for the thin account of agency, since it is one of the reasons why the thin account

³⁰See List & Pettit (2011, 35).

³¹Pettit (2017, 18)

³²List & Pettit (2011, footnote 64)

falls prey to the attributionist challenge. I will raise that challenge at the end of this chapter. Furthermore, for the thin account, there is no inconsistency in interpreting aggregation functions (organisational structures) as shared policies of procedure, because we can easily think of these procedures as specifying concrete actions as inputs. Their intentional nature has no bearing on what they specify.

It is also worth noting how L&P characterise the relationship between aggregation functions and the group at different points in their book. Often, it is not members of a group who form a shared policy of procedure. Rather, the group itself implements a certain aggregation function by introducing a new organizational structure. This does not mean that we must abandon our interpretation of aggregation mechanisms via shared policies of procedure. It means only that the formation of shared policies is itself direct and governed by the group as a locus of agential control. But this is not necessarily the case. For groups such as public corporations, characteristic organisational structures – like governance by a board, for example – are externally imposed and required by law and cannot be changed by the group itself.

L&P also allow for groups which non-intentionally aggregate attitudes.

Individuals may form a group agent in virtue of evolutionary selection or cleverly designed incentives to act as required for group agency, perhaps within independent cells. Here the individuals contribute to the group agent's performance, but do not explicitly authorize the group agent; they need not even be aware of its existence.

The groups referred to here seem to be more like swarms of bees or ants, unlikely to display more sophisticated agency. This is consistent with L&P's beliefs, which are revealed when they consider Block's famous thought experiment of the China-body. This body replicates a person's mental life by assigning 'mechanical' tasks to each member of the Chinese population, thereby creating a computer-like system which manifests the same kind of functional properties as the mind of the individual which it is replicating. They note that, to output the group's attitudes, the aggregation function clearly takes the individuals' non-attitudinal contribution into account. But they also think that such a possibility is highly unlikely. Therefore, they conclude that aggregation functions must supervene on the attitudinal contributions of the group members. I will follow them on this issue, when I later present my own account insofar as I think that robust group agency requires relevant group mental states.

Critical Thoughts About Aggregation functions

At this point, I want to highlight a tension in L&P's theory of group agency. As I have already mentioned in Chapter 1 on the intentional stance, L&P claim that attitude attribution should be driven by a group's behaviour. In their account, the analysis of aggregation functions plays several roles. On one hand, they show that our assertion of the rationality of group agents can be justified by appeal to the kind of aggregation mechanisms used by the group. Thus, certain types of aggregation mechanisms – like the premise-based procedure or some kind of collective deliberation – provide stronger evidence for the thesis that the group is rational, while other aggregation mechanisms provide less evidence. With regard to other standards of rationality, for example, whether group beliefs accurately track the truth, L&P again show that certain aggregation functions more clearly indicate that the group meets the standard of rationality by having beliefs which aim at truth. The tension comes from another conclusion about aggregation functions. On one hand, the authors claim that aggregation functions may be read off the organisational structure which is introduced via shared policies of procedure. But, as they say later on, this need not be the case. We may also interpret a structure that has come about non-intentionally as an aggregation function. But L&P do not specify how to do this. We may suppose that, in this non-intentional case, we simply make use of the methodology of the intentional stance and find an intentional profile that fits the behaviour. This is not always easy, since, often, more than one intentional profile might explain the same type of behaviour. Bringing aggregation functions into the picture confronts us with the same problem, but at a lower level. Each intentional profile might be realised via different aggregation functions. Further, since the resulting group attitude may be such that it is not shared by any other member, it seems strange that we should be committed to considering only the individual members' contributions. At least in analysing and explaining the group's behaviour, appeals to members' contributions, given complex enough aggregation mechanisms might not help us in determining the true group attitude.

2.4.3 Group Reasoning

Now, we are ready to answer the question that we raised in Section 2.4. How can a group reason and deliberate in order to determine, for example, which of its beliefs or desires should function as premises or as a conclusion in a premise-based aggregation function?

Let us begin with some intuitions. Suppose that we are a research group

interested in determining which books to read in the upcoming semester. One way for us to proceed is to simply let the most senior member determine which book to read, having informed her of our interests. Alternatively, we may determine what to read by asking ourselves what would best advance the research of all the group's members. Both of these modes of shared deliberation are nothing like the mechanical majority vote that our board of directors used on the fracking question back in Table 2.1.

The difference between such an aggregation function and a more formal one – such as the British people's direct democratic vote on the question of whether to remain in the EU or to leave it – is that the latter is *functionally explicit*, whereas the former is *functionally implicit*. It is functionally explicit in that the condition under which the British people count as having decided against remaining in the EU are precisely defined. On the other hand, in deliberating about something like a list of readings, the participants of the deliberation will have in mind no mechanically-enacted aggregation function such as in the case of a referendum to be settled by a simple majority. This permits a deliberative process in which, due to deliberation, the participants' preferences or beliefs change in favour of the group perspective. Nevertheless, List and Pettit maintain that even for these complex processes, an aggregation function can be found which need not be employed again.³³

I have just mentioned one very important aspect of deliberation and of the functioning of a group: in considering what beliefs or goals the group ought to adopt, it is not always possible to achieve group rationality by simply letting each member consult their own preferences or beliefs about the matter. This individual consultation is not helpful because, in discursive dilemma situations, diverging individual attitudes will lead to inconsistencies on the group level. In my discussion of discursive dilemmas, I assumed that the company's board members were unaware, and could not become aware of each other's beliefs. This assumption is plausible if the board members did not deliberate before voting and cast their votes anonymously. If, on the other hand, the board members did talk, deliberate together or know each others preferences before the vote, they might have been able to anticipate a pattern of a discursive dilemma. Thus, they might form the belief that, if they were to vote in a particular way, this would lead to an inconsistency among the group's attitudes, along with the corollary belief that, if they were to vote differently, voting differently would lead to group-level consistency. Desiring that the group's attitudes remain consistent, they might change their personal attitudes accordingly. This anticipation of patterns on the collective level, and use of this information to adapt one's own individual attitudes or

³³List & Pettit (2011, 63)

input before they are fed into the aggregation mechanism, is what (List & Pettit 2011, 62-63) call *feedback*. Feedback is a key mechanism in explaining how group deliberation can impact the members' individual attitudes so that they are more in line with the group's overall point of view, consisting of the group's beliefs, goals and intentions.

Feedback mechanisms also play an important role. They help the group to form relevant meta-beliefs when one of its members realises that the group has gone astray. Thereby, feedback mechanisms are a necessary element in a thin form of *collective reasoning*.³⁴

Let's consider this claim more closely. Suppose that an individual agent who believes that *p* and believes that if *p*, then *q*, asks herself whether *q*. Now, the agent will form a meta-belief with the following content: from the two premises *p* and if *p*, then *q*, it is valid to infer that *q*. This meta-belief about the validity of an inference, together with the desire to be rational imposes a check on the rational process and will lead the agent to form the belief that *q*. The meta-belief allows her to intentionally reinforce her own rationality.³⁵ Not all meta-beliefs need be concerned with the validity of an inference. The agent may also ask herself whether a given combination of propositions is consistent or well-supported by evidence or whether a given proposition is desirable or good. Intentionally forming these metalanguage beliefs about propositions in order to impose checks on rational processing is what List and Pettit understand as 'reasoning'.

How might such reasoning come about in group agents? First, we may note that within List and Pettit's framework, there is no conceptual limitation to an agenda containing beliefs about propositions such as the belief whether two propositions are consistent or inconsistent. Nor is there any general limitation to the aggregation function: it may take as input not just the individual's propositional beliefs, but also beliefs about a proposition, for example, whether it is consistent. Here, feedback mechanisms come in. Feedback allows each individual agent to form beliefs about the group's attitudes. Then, if one agent makes public such a belief – for example, that the group attitudes in a discursive dilemma with majority voting turn out to be inconsistent – then the group can collectively form the meta-belief that voting only on the premises restores consistency. Continuing with this scenario, each member can convince herself of the truth of the meta-belief in question by considering the group's attitudes. Thus, via feedback mechanisms, each member of the group can form the relevant belief.

³⁴Although List & Pettit (2011, 64) don't commit to this claim, they say as much when they discuss different organizational structures of groups.

³⁵List & Pettit (2011, 30)

This is the overall structure of theoretical and practical group reasoning according to the picture of agency which corresponds to the thin account of personhood. On this view, forming meta-beliefs about how one's attitudes hang together, such as whether they are consistent or whether they are akratic (that is, that they are situations in which the group believes that it ought to X, while not intending to X). This kind of reasoning requires policies of procedure that allow room for feedback mechanisms and for functionally inexplicit aggregation functions that take as input judgements on normative propositions. Furthermore, it requires that the agent can act on these beliefs. This brings us to the question of what it means to act for a group under the thin account of personhood.

2.4.4 Group Action Under the Thin Account

At first glance, it is somewhat surprising that List and Pettit's theory of group agency is not explicitly committed to an account of group action. Nevertheless, from what they say in passing, we may assume that they have in mind a Davidsonian account, at least with regard to their thin account of agency.

It [the agent] has the capacity to process its representational and motivational states, leading it to intervene suitably in the environment whenever that environment fails to match a motivating specification.

[...]

When action is taken in the group's name - say by its members or deputies - this is done for the satisfaction of the group's desires, and according to the group's beliefs.³⁶

Their emphasis on beliefs and motivational states that explain action and intention is clearly in the spirit of a Davidsonian account – an account which reduces intentions to a combination of beliefs and desires. Although L&P mention that human action should be understood via Bratman's planning theory of intentions,³⁷ they do not suggest that this is how we should understand the actions of groups.

For this reason, I take L&P's thin account of agency to correspond roughly to a Davidsonian understanding of action. Later on, I will discuss their account of agency connected to a thick account of a person, where group (self-)identification plays an important role and shared policies of giving weight

³⁶List & Pettit (2011, 20 and 32)

³⁷List & Pettit (2011, 22)

to different values enter the picture. In that case, the resulting account of action approaches an understanding of will more closely connected to deliberation and reasoning than on the present picture, which requires only that the group is able to act on normative judgements, but does not closely tie those judgments to deliberation. On the present account, it is sufficient to form just a normative judgement – say by voting on the conclusion of each persons’ individual deliberation. Thick accounts, on the other hand, place more weight on the will as ‘over and above’ the operation of event-causal factors [...] and as accessible only from within the standpoint of deliberation’.³⁸ On the thin account, group deliberation is not necessary for forming normative judgements. The resulting intentions of the group, which can be reduced to combinations of beliefs and motivational states, therefore do not feature in the deliberation of the group as they would were they irreducible or primitive volitional states, like an intention in Bratman’s plan-theoretic understanding of the will.³⁹ Thus, a group which only satisfies the thin account can only form intentions which do not constitute what we would typically understand as a will, because those intentions lack the volitional character that is normally associated with will.

Enacting Group Intentions

An important further part of List and Pettit’s account of action is that the relevant belief and desire pair leads to a group action, such that the group’s organisational structure both ensures that enactors carry out the relevant action and that substitutes are found in case the original members are not available.⁴⁰

Since group action involves enactors being controlled by group motivational specifications and group beliefs, L&P argue that it is important that organisational structures be compatible with incentives. In other words, a group must ensure that the enactors’ desires align with those of the group. More precisely, a group agent’s follow-through with its intention is *incentive compatible* if it is rational for the individuals to do their part in satisfying that group intention.⁴¹ The need for an organisational structure that is incentive compatible is especially urgent in those cases in which a group attitude is not shared by its members – for example, a public corporation aiming for profits, when the managers would rather just fill their own pockets. The absence of an organisational structure that is incentive compatible will lead

³⁸See Bratman (1999, 276).

³⁹Bratman (1999, Chapter 11)

⁴⁰List & Pettit (2011, 37).

⁴¹List & Pettit (2011, 106 and 123)

to weak-willed groups.⁴²

Incentive compatibility can be achieved by either changing the organisational structure so as to make strategic, non-truthful voting individually irrational, without trying to influence the members' existing attitudes and dispositions. This change can be achieved by incentivising certain options and strategies, while disincentivising others. It can also be achieved by changing the preferences and attitudes of the members through persuasion, argumentation or another form of team-reasoning, we-thinking or shared intentionality about the group action.⁴³

So, on the thin account, there are two rational dimensions to aggregation functions which issue in actions. On one hand, the outputs of these aggregation mechanisms must satisfy attitude-to-attitude standards of rationality: decisions must be consistent with each other and the group must intend the means to its ends. On the other hand, aggregation mechanisms must be incentive compatible to meet standards of rationality that concern attitude-to-action relations and rule out weakness of the will. This concludes my exposition of the thin account of group agency.

2.5 The Thick Account of Group Agency

L&P distinguish the thick account of personhood or group agency from the thin account by means of the following characteristic form of reasoning: '[t]he special reasoning that invokes the very concept of a reason, registering that the truth of some propositions provides a reason *for me* to hold by another proposition, may require self-identification.'

⁴⁴ They claim that for a group to act for a reason, the group must self-identify. This moves their account action by a person in this thick sense closer to a theory of the will, rather than a mere event-causal story.

2.5.1 Group Identification

L&P begin the construction of their notion of group self-identification by analysing the more basic notion of group identification. Group identification occurs in response to an identification gap. For group agents, an identification gap arises when one of the group's desires or attitudes involving the

⁴²Pettit (2003) calls this corporate akrasia.

⁴³For different accounts of team reasoning, see Bacharach et al. (2006), Sugden (2000), Gold & Sugden (2007) and Tuomela (1995).

⁴⁴See List & Pettit (2011, 228), although this claim is defended by Pettit alone. As I have said before, List thinks that the thin account of group agency is, by itself, sufficient.

group under some neutral description fails to activate the group because the group members or enactors fail to recognise themselves as the referents of the description.⁴⁵ According to the authors, the gap is a failure of instrumental rationality. For example, a desire represented in the propositional mode, such as ‘The book club that meets in room 105 ought to clean that room’ fails to lead to the instrumentally required infinitival expression ‘We ought to clean room 105’ because each member, as a member of the group, fails to realise that they constitute the book club that is supposed to clean the room. This might be because they all belong to several book clubs and fail to recognise the description’s referent as the group to which they belong. It could also happen that they simply forget that they inhabit the relevant role, being the member of the book club whose responsibility it is to clean room 105.

Identification gaps share certain features with weak-willed corporations. A corporation is weak willed when it forms an intention through some organisational structure, but this intention fails to activate the relevant enactors because the organisational structure is not incentive compatible. The enactors are simply not motivated and not moved by the corporate intention. In these cases, incentive compatibility can be achieved either by introducing another decision mechanism which is incentive compatible – perhaps one in which the decision is always in line with the individual’s attitudes (a conclusion based decision procedure) – by introducing monetary incentives which transform the members’ preferences to bring them in line with those of the group or by motivating an independent affection for the group on the part of the members.⁴⁶

The identification gap may also be closed if the members adopt the group’s perspective or point of view. A member who adopts that point of view aligns her attitudes with those of the group by simply adopting the group’s attitudes and automatically acts out what her role requires of her.

L&P emphasise that, when the individual members of the group adopts the point of view of the group and lets herself be guided by it, then “we, the members, each have attitudes in whose propositional expression the group figures as we.”⁴⁷

In adopting the group’s point of view, members ‘may connect themselves directly to the attitudes of the group. They do not treat the group attitudes as mere indicators of what the group is to do, asking themselves explicitly

⁴⁵See List & Pettit (2011, 187) for an individual account of the same phenomenon and List & Pettit (2011, 191) for the collective version.

⁴⁶This affection is extensively discussed by L&P under the concept of becoming ‘invested’ in group agents. I will explore it in further depth in the next section on group self-identification.

⁴⁷List & Pettit (2011, 192)

whether they wish to identify with the group, and acting only if they have this wish. Rather, their individual attitudes are under the automatic guidance of the group [...] Or at least they may do this when there are no 'red lights' that suggest they should hesitate and take stock.'⁴⁸ L&P describe this phenomenon as the group's mind being instantiated in each of the members. The members *identify* with the group, and it is this kind of identification by means of adopting the point of view of the group which, produces the more complex phenomenon of group self-identification. The thin account differs insofar as it does not specify how the identification gap is bridged, that is, whether it is closed by mere incentive compatible aggregation functions. An incentive- compatible identification function does not yet ensure that the members have attitudes in whose propositional content the subject is specified as 'we'.

This account of adopting the group's point of view resembles Bacharach's 2006 account of team reasoning. On that account, people spontaneously team reason in order to coordinate their actions.⁴⁹ Bacharach argues that when a group of agents reasons as a team, each member frames decision situation from the perspective of the group. Framing a decision situation in this way entails an agency transformation, from an 'I' to a 'we'. This transformation has two functional aspects. First, members consider the *choices that are open to the group* instead of the choices that are open to them as individual agents. Second, they rank these group choices according to the *group's preferences* instead of their own. Finally, the individuals do their part in the group's action. I want to draw attention to this parallel with Bacharach's theory, because Bacharach assumes that what is required for this process demands that all the members agree on the choices open to the group and on the group's preference ranking, where both of these distributive features of team reasoning are common knowledge. In the case of group agents that are constituted by organisational structures, this agreement is ensured by the shared policies of procedure. Because the relevant members share these policies, their output is also collectively accepted. This output, again, determines the elements which make up the group's point of view and ensures that, in an ideal setting, the members all agree about the group's ends of the group, how the group represents its environment and the options open to it.

⁴⁸List & Pettit (2011, 192)

⁴⁹See Tuomela (2013) for a philosophical account of team reasoning that is influenced by Bacharach's approach and Sugden (2000) and Gold & Sugden (2007) for variations and different applications of Bacharach's concept of team reasoning.

2.5.2 From Group Identification to Self-Identification

As I've described, when members of a group identify with it by means of adopting its point of view, they will refer to the group's attitudes as 'their' attitudes. They will employ the first-person plural pronoun which allows them in a linguistically mediated way to *self-identify* with the group. Importantly, although they will talk about their attitudes, the referent of the 'we' is not the sum of the group's members: it is, but the single, unified group agent, the center of agency which the members constitute. The group's self counts as 'us' or as the group's metaphysical self in virtue of moving those members who adopt the group's point of view.

Two Features of Group Self-Identification

Self-identification, according to L&P, has two features.⁵⁰ First, it allows the agent to pick out their self as one among other selves, which can stand in reciprocal relationships to other agents. This entails identifying the practical identity of the group which distinguishes it from others and by which the group is recognizable. We I will say more about the practical identity of a group and what else it is constituted by, when we come to the concepts of owning and disowning that shapes a group's character or its practical identity (Section 2.5.2).

Second, by virtue of self-identification, the agent enters into a special relation towards its own mental states: a *special authoritative relationship* that allows the members to use the indexical 'we' and to speak for the group.⁵¹

Earlier, I have said that the identification gap is closed if the group members adopt the point of view of the group. But, so far, this says nothing about why they should do this or why they should be motivated by the ends thereby adopted. Members' being motivated by the point of view and acting accordingly requires that the group agent '[makes] the same intimate claims on them as their individual selves do.'⁵² This is because group members can affiliate with the group's self in the same way as they affiliate with their individual selves. As List & Pettit (2011, 196) put it, people '[...] may find a recognizable self not just in the individual persons they are but in any group person they are prepared to serve.' L&P call this process *investing* in group agents. People can invest in their personal self, but can also favour a group's self instead, choosing to be prompted by the group's point of view

⁵⁰List & Pettit (2011, 188–190, 194)

⁵¹See also Pettit (2017).

⁵²List & Pettit (2011, 196)

rather than their own point of view.⁵³ An examples of this kind is the abbot Abbot Narcissus from Herman Hesse's 'Narcissus and Goldmund'. Narcissus' life is the church, which moves him in thought and action and to which he dedicates his life. The notion of being invested in a group is affective in nature and certain group's are more likely to produce this relation than others. An employee member of a corporation which is largely profit- driven will be unlikely to identify with the group in a way such that the group's attitudes make the same intimate claim on him as his individual self does. In those cases, identification will typically be achieved via incentive mechanisms that do not lead to L&P's notion of self-identification.

The combination of being invested in a group agent and being aware of it as one agent among many carries with it an aspect that L&P call *amour propre*. Amour propre consists in passions 'prompted by the recognition among individuals that a group agent with which they self-identify exists in competition with others. [...] What this sentiment involves is a desire, recognized and authorized in the group, to perform on a par with competing entities or to outperform them.'⁵⁴ L&P argue that it is amour propre which gives rise to shame and humiliation when the group agent fails to meet expectations or pride and enjoyment in group triumph.

Owing and Disowning

Since the particular identity that might be associated with a group agent can have the powerful emotional effects as captured with the term amour propre, a group, just like an individual, might want to disown certain past actions performed which are cause for collective grievance and own other actions of which the collective is proud as representative of actions that speak for the group's identity. This process of owning and disowning certain actions carves out a group's character. It is not only driven and explained by the individuals' wish to be associated with a group of whose achievements they can be proud, but also a matter of the group being a potential partner in different kinds of relationships involving trust and a good will –, since a past pattern of bad actions, indicative of a bad corporate character, might undermine those features of the group which are necessary for the group to be trusted in certain matters. Thus, a group, like one-time Nazi collaborator Volkswagen, has worked with ill-reputed regimes, might want to disown this episode if only to avoid being disqualified from future collaborative ventures. It is in a group's interest to present the best possible version of itself.

⁵³That individual and group selves can enter into such a conflict is also discussed by Rovane (1997, 140–141).

⁵⁴List & Pettit (2011, 190)

When members disagree about which actions should be owned and which disowned, there can be a ‘struggle for the ‘soul’ of the group. Different members may fight for rival visions of the character of the group in which they are invested. When this happens, the construction of the group’s character will take on a directly narrative form, with rival members telling different stories of where the group is coming from, and what episodes characterize it best.’⁵⁵ The procedure of Owning and disowning targets not only episodes from a group’s past and what role they play to justify future policies, but also values more generally.

Although L&P do not mention it, the result of such struggles can – uniquely in the case of group agents, but not individuals – be the fracturing of the group into the original group that adheres to policies consistent with its past and a revolutionary movement which forms its own group as a result of unsuccessful attempts to move the group’s character or standpoint.⁵⁶

Including the mechanism of owning and disowning into the picture of self-identification adds greater complexities to L&P’s view. Different subsets of members may self-identify with the group in different ways such that they interpret the point of view that they attach to the group differently. Since the point of view is necessarily incomplete, different interpretations lead to different conclusions on normative issues. Suppose, for example, that a corporation in the mining business has adopted the vague end of attaching importance to cultural artefacts and sites nearby their drilling operations, but has also adopted the end of maximising profits. Different parts of the group will have various ways of understanding the constraints that these ends place on each other in different ways. Some might understand the end as prohibiting drilling in locations which are sacred sites, while others will interpret the end as relevant insofar the concerns about the desecration of holy sites are raised in a public manner thereby threatening the corporation’s reputation and, indirectly, its profitability. Clearly, if the point of view is interpreted in these different ways, then certain interpretations will move the members while others will not move them, which can cause these members to feel alienated from the group and incapable of self-identifying with it. Members feel alienated if the group acts in a way which is incoherent with their interpretation of its practical identity. Self-identification is shaped by the group’s internal structure and the policies of procedure of the group, which means that not all group members will have an equal role and influence in shaping its point of view and, but also in owning or disowning its previous

⁵⁵List & Pettit (2011, 201).

⁵⁶A case in point is the German far left party ‘Die Linke’. The party, which emerged after internal dispute within the SPD, which was, at the time, the main left-leaning parliamentary party.

actions and values.

Before we summarise this discussion, note that the mechanism of owning and disowning introduces narrative elements into the picture of group agency that might transcend a non-cognitivist picture of agency, suggesting, instead a cognitivist framework of agency. Such a framework, in which agency aims constitutively at self-intelligibility, has been developed by Velleman. In this dissertation, my claim is that such a framework is better equipped to counter the challenges to L&P's account that I will discuss in the next chapter. I will also show that it can incorporate much of L&P's framework.

We can now try to give a concise definition of the account of group self-identification.

Definition 2.5.1. *Group Self-Identification:*

A group agent G self-identifies as G iff its members identify with G such that they take up G 's point of view, which allows them to speak and think of the group using the indexical 'we' such that

- (a) they can **authoritatively** speak for the group*
- (b) they are **aware** of their group and their group's identity as one among many groups*
- (c) if identification happens, because they are **invested** in the group, then b) leads them to feel shame and pride on the group's behalf*
- (d) (c) gives rise to the members participate in sculpting a **shared narrative** of the group's practical identity by collectively owning past actions that make them feel proud and disowning past events that make them feel ashamed.*

In this definition, the notion of a group's practical identity appears. The practical identity is not to be confused with the group's metaphysical identity; but it is that part of its identity by which the group can be picked out among other groups and by which it can be described and distinguished. I will say more about this in discussing my constitutive account of agency in Chapter 5.

A unique feature of change by the notion of self-identification compared to mere identification with the group is not only how the group's point of view relates to its members, but also how that point of view is constituted. Not only can members authoritatively speak for the group, because they identify with it. Their relation of identification can be such that the group's self makes the same 'intimate' claims on them as their individual self does. For individuals, this relationship of being invested in the group, can produce

feelings of shame and or guilt on behalf of the group. The relation also adds new states to the point of view, namely the past events and values which are owned or disowned.⁵⁷ The dynamics of owing or disowning some and disowning other states are fuelled by the feelings of shame, humiliation or pride on behalf of the group. They also and give rise to a group narrative. But this introduces an element into L&P's account which makes it more difficult to conduct analyses via by means of aggregation functions whose input are manifested attitudes.

Group Self-Identification vs Plural Self-Awareness

L&P's theory of group self-identification shows quite an overlap with the concept of plural (pre-reflective) self-awareness developed by Schmid (2009, 2014). Just as self-identification gives rise or constitutes the group's self, so does plural self-awareness, according to (Schmid 2014, 14), "unify the mind and constitute selfhood." According to Schmid plural self-awareness, which he takes to be the prerequisite or background condition for groups of people to act together and as such has a much more broader scope than L&P's concept of self-identification, has three characteristic features to it.

First, it involves a *normative* stance towards oneself or, in this case, the group's mind, that gives rise to intentions' normative force and the truth tracking of beliefs. Second, it involves assuming the *perspective* of the group, which entails the ability to self-reflect and see oneself as one agent among others, thereby constituting a group's perception of its own practical identity. Third, it involves *ownership* of the group's beliefs, intentions and goals as our own. All of these aspects also feature in L&P's account of self-identification, although, in Schmid's account, the normative force seems to apply to all members equally. This normative force is limited in L&P's case by the hierarchical structure of groups: I can only resolve inconsistencies that are up to me (or us) as specified by the shared policies of procedures that constitute my (our) group. Part of this difference in normative scope is due in part to the disparate phenomenon both accounts seek to explain. Schmid is interested in small, democratic groups acting together without hierarchical structures. L&P, on the other hand, give an account of highly structured groups in which members might not know about each other or and about most other mental states of the group. Nevertheless, it is important to highlight the similarities, because plural self-awareness is an essential part of cognitive theories of actions. The constitutive account of group agency, inspired by Velleman's cognitive account of action, is one species of that genus.

⁵⁷List & Pettit (2011, 200–201)

Self-Identification and Sufficiency

The mechanism of owning and disowning is shaped by the organisational structure. But once we have established which attitudes are owned by the group and how they fit into the corporate narrative, we might inquire what the rational interrelations between the narrative that constitutes the group and the mental states of the group agent are. As I have argued above, different interpretations of the group's character or the group narrative can have an influence over what the group should do and what values it should adopt. In L&P's terms, different narratives make it easier for members to 'argue' for certain policies. 'If the United States is a fundamentally secular state,' they maintain, 'created with a conscious distancing from religious faith, then it is easier for liberals to argue against any favoritism towards the Christian or Judaeo-Christian tradition.'⁵⁸ These observations take us in a direction where group action is not merely a combination of a belief-desire pair together with a controlling mechanism which realises this state – like an incentive compatible organisation structure. We are now talking about policies or belief-desire pairs that a group can self-identify with and thereby own or disown. Here, self-identification involves a shared identity, a character that rationalises certain actions and makes them speak for the group. This seems to extend the concept of a group's point of view, which we previously understood only as the set of attitudes that are generated by the group's aggregation functions. Under the understanding of self-identification we have arrived at now, the point of view is not just simply generated by any aggregation function, which includes a decision mechanism and a (set of) enactors sufficiently motivated to realise it. Instead, these enactors as well as those individuals making the decision, must be motivated so as to adopt the point of view of the group, because of an affiliation with it – , which we understand as being invested in the group. Finally, the group reason must cohere with the group's character such that the group owns the resulting group action.

There is a problem with this way of understanding self-identification. First, it seems to tie self-identification too closely to a particular kind of emotional relation, which does not obtain within many groups, which seem to be able to act intentionally. Many firms seem to have the functional equivalent of self-identification, while lacking the emotional bond between the members and the group. At the same time, some members have an overly strong attachment to the group, while the group lacks the kind of rationally unified point of view that is required for agency. Football clubs are a good example of this kind. This means that we should understand

⁵⁸List & Pettit (2011, 201)

group self-identification as at best a *sufficient* condition on realizing proper group actions, but not a necessary condition on full-fledged group agency. Later on, I will challenge this claim, when I discuss an attribution challenge for proceduralist accounts in Chapter 3.

2.5.3 The Thick Account of Agency and the Group Will

To what extent does group self-identification ensure that the group acts for a will, rather than due to the causal efficacy of a belief-desire pair?

Self-identification adds more to the picture than only the specific kind of reasoning that consists in framing the question of what the group ought to do as a question of what *we* ought to do. It adds a further element to the thin Davidsonian account of action. By requiring that the members who, in self-identifying with the group, are emotionally and affectively invested in the group meaning that they ‘own’ and ‘disown’ certain group actions for which they in turn can feel pride and shame, L&P seem to introduce a kind of endorsement of first-order desires. This changes the account into a Frankfurtian hierarchical model of action. As Fischer (2012) has argued, this might already be enough for a weak account of a will if we understand the second-order endorsement as a second-order volition.⁵⁹ But it is not clear that this is true. Watson (1987) has argued that not any second-order desire or endorsement can, by itself, ensure that the resulting action of an intentional system is one that the group identifies with and is therefore a fully intentional action. This is especially so, since there might be nothing special about the second-order desire that lends it special authority as being expressive of the agent’s self, such that the agent is fully responsible for it.

The will, as Bratman has argued, has both a volitional and a reasoning-centered aspect.⁶⁰ The volitional centered aspect explains how intentions or the state that is identified as an intention determine what we do, whereas the reasoning-centered aspect influences and restricts both the scope of our deliberation and provides an input into our reasoning. On this second aspect, the self-identification account of agency remains exceedingly vague. It does specify the conative nature that underlies the idea that a group owns an action, by pointing at carrying out an action because of being invested in the group additional to the existence of incentive compatible mechanisms. But, it fails to specify the role that these intentions or belief-desire pairs that group members act upon, when they group self-identify, play in further reasoning. Simply framing a decision as ‘*we* have previously decided to X

⁵⁹Frankfurt (2018, 84)

⁶⁰Here I follow Bratman’s account by Velleman (2012).

that *we* own' doesn't tell you the role that this intention or intentional state should play. Suppose it comes into conflict with another intention: should the intention be reconsidered, or should it remain?. If this decision is up to me, given my position in the hierarchy, then knowing only that *we* have formed this decision previously, tells me nothing about how I should interpret it with regard to further means or conflicting intentions. Consequently, this means that the intentions' volitional aspect is also threatened if I have no principled basis on which to resolve future conflicts among other intentions or if group self-identification just does not guide my other practical reasoning on behalf of the group.

2.5.4 Self-Identification and Collective Reasoning

Although self-identification is one way of making sense of modes of reasoning that conclude with an intention – enkratic and instrumental reasoning – it does not exhaust the normative reasoning of which a moral agent must be capable of. Any more sophisticated agent is faced with conflicts among their normative ends and must systematically adjudicate these conflicts. As Korsgaard (2003) argues, this is constitutive of an agent who believes that we have to adopt and commit to principles of choice so as not to be a wanton – that is moved in this direction in one moment, just to be moved in the opposite direction in the next minute. As I argue in Section 5.5.1),⁶¹ avoiding this kind of diachronic irrationality is constitutive of what it is for an agent to have a will that is capable of playing the characteristic functional roles of settling practical issues for the agent and of resolving (incommensurable) conflicts among one's ends. Working out which options are open to me, resolving conflicts, ranking options according to different preferences and assessing the moral consequences and moral features of one's actions are all different parts of deliberative processes.⁶² They are typically required for working out one's beliefs about what one ought to do. Capacities for these deliberative processes across time cannot just be explained simply by appealing to reasoning via meta-beliefs, nor by appealing to self-identification, since nothing in the concept of self-identification and reasoning via meta-beliefs restricts the group agent from repeatedly changing its mind about how it ought to rank various options. Perhaps the agent even believes that it faces an incommensurable choice, because the rationales of its members are divided between two normative perspectives which stand in reflective equilibrium. This does not mean that there isn't a way to build this diachronic

⁶¹See Altner (2021).

⁶²See Rovane (1997, 21ff).

deliberative mechanism into L&P's theory. However, it's just that in their 2011 work and in Pettit (2017), they do not present tools that are fit for the job.

This gap in the account of how group agents come to *form* moral judgements is also noted in Hindriks (2014, 2018), who understands it as a failure to provide an account of 'normative competence' for group agents.⁶³ Hindriks notes that a collective in which members only 'infrequently and [on an] *ad hoc* basis' contribute normative considerations to decision-making which gives rise to group attitudes do not count as having normative competence.⁶⁴ The idea is similar to Millikan's condition for applying the intentional stance that we discussed in Section 1.5.1. She argued that we should not adopt the intentional stance towards a system and on its basis predict future behavior, if its rationality was due to chance, rather than robust dispositions and a sensible design. Similarly, we should not view a group as an agent capable of moral judgements, if the formation of previous judgements was purely coincidental.

In what follows, I will introduce Bratman's account of shared deliberation, which Hindriks uses to build an account of a normative group perspective. After introducing this account, we will have all the pieces in place to produce the strongest possible proceduralist model of a moral group agent. As it turns out, this model will require further amendments.

2.5.5 Bratman's Account of Shared Deliberation

Central to Bratman's concept of shared deliberation is the idea that, in deliberating, we appeal to a *common ground of shared values*. Values that we are all committed to regarding as important or mattering or being relevant in our deliberation. We can understand this as a commitment to *giving weight* to certain kinds of considerations in our collective reasoning.

Commitments to weights can't be held individually for them to function as a common ground. They are neither sufficient nor necessary. Consider the following argument by Bratman.

Perhaps each of us thinks earthquake safety in construction is a good thing and these judgments are common knowledge, yet we still do not have a shared commitment to give weight to earthquake safety in our shared deliberations concerning our construction project. Perhaps some of us resist such a commitment be-

⁶³We might also think of the capacities that give rise to an agent's normative competence as its normative profile, but I will return to this issue in Chapter 5.

⁶⁴Hindriks (2018, 7)

cause, though earthquake safety is a good thing, it would stand in the way of higher profits, or speed of construction.

[..]

public convergence of value judgment is also not necessary for a corresponding shared commitment to weights. We could have a shared commitment to, for example, our giving weight to collegiality in our hiring decisions, or to legacies in admissions decisions, even if some of us are to some extent skeptical about the associated values but participate in the shared commitment as a part of a social compromise.⁶⁵

Commitments to weights that constitute a common ground must therefore build in two ways on the notion shared intentionality that we have discussed previously in Section 2.4.1). On the one hand shared deliberation happens within the context of a shared action itself. We might deliberate on what and when to publish as part of our shared activity to do research together or we might deliberate how to produce, promote and sell a product as part of our shared activity to make a profit by selling this product. On the other hand shared deliberation is itself a shared activity. It involves the kind of interlocking, interdependence and mesh that is characteristic of the idea of shared intentionality.

Shared commitments to weights not only support mesh between sub-plans in situations of coordination and cooperation, they are also definitive of certain shared activities and help to unite the members that partake in that activity. Bratman gives the following example.

If you are going to be a member of a certain scientific research team, you may need to participate in a shared commitment to giving weight, in discussions of the direction of the research, to the potential for lucrative patents. Central to certain groups may be a shared commitment to giving weight to conformity to particular religious texts or traditions or rituals. Members of a club might have a shared commitment to giving weight to religious affiliation, or gender or race, in shared deliberation about membership; and that may be why you do not want to be a member of that club.⁶⁶

Shared intentions about weights can only reliably and predictably guide the reasoning of the members of a group, and can only help the inputs of team members in shared reasoning be explanatory and intelligible over time

⁶⁵Bratman (2013, 136-137)

⁶⁶Bratman (2013, 135)

if the group can't form a new shared intention about weights every time it deliberates. Agreeing on what the group commits to count as normatively weighty every time it meets is not only costly, but it will also inadvertently lead to inconsistencies over time. Therefore, shared intentions about weights, which structure our shared reasoning, must be sufficiently general, so that they to apply to similar decision situations across time. These shared intentions about weights will, therefore, in diachronically extended activities, like the activities of groups, be *shared policies* about weights. These policies establish what considerations members assign normative weight to in situations of a certain kind. A company might have the a policy to assigning weight to environmental considerations whenever these might be of concern to the local community or its costumer, coupled with a complementary policy to ignore environmental issues otherwise. These policies are not absolute, but rather defeasible, given that there is the presence of agreement about what counts as a defeating condition. The corporation might have a policy to pursue strategies that maximise profits unless it can be expected that these policies will cause long term losses.

As I have already mentioned, one reason for why groups adopt shared policies about weights rather than mere shared intentions about them, is the need for stable commitments towards weights over time as a means of ensuring relevant cross-temporal coordination, cooperation and adjudication of conflicts among the group's competing goals. What does it take for policies about weights to persist over time? For shared intentions to persist, they must be relevantly supported by reasons concerning either the feasibility, the desirability or the obligatoriness of the shared intention. When these reasons persist over time, so do the shared policies concerning weights.

It is important to note that these reasons may differ across group members. Take a company's shared policy to make profits by selling vaccines. This policy might be supported by those members, whose pay is connected to the overall profits and so their reason for participating in this policy is due to their desire for money. Other members might have promised the companies owners to act in their interests, who want the company to make a profit – they have an obligation-based reason to participate in the policy. Finally, others might simply participate in the production and sale of vaccines and, given that the only feasible way for the company to afford that activity it needs is to profit off of vaccine sales. They participate for feasibility-based reasons. This permissibility concerning individually diverging reasons for participating in a joint action is a key characteristic which distinguishes the Bratmanian framework from alternatives such as the subject account endorsed by Gilbert (1987, 1990).

For shared policies concerning weights to resolve conflicts between nor-

mative considerations – such as profit-maximizing versus environmental concerns – in addition to enabling the group to agree on what reasons to appeal to in shared deliberation, these shared policies about weights must have to assign a *relative* weight to the different competing considerations, such that this relative weight is *accepted* as part of a shared policy. Shared policies of acceptance with regard to weights allows the group to reliably, predictably and coherently resolve conflicts over time. Acceptance allows the participants in shared deliberation to take what is accepted as *given* in deliberation. We may visualise this web of shared, accepted policies about relative weights as inducing a *ranking* between different goals in the situations which those policies target. Bratman stresses that this ranking is only partial and does not fix every aspect of the decision situation. Rather, as I have said before, it provides a common ground which guides the group’s reasoning and joint deliberation. To resolve conflicts among values that are assigned similar weights, policies must have a certain depth. Conversely, diachronically coherent resolutions will be less likely when the relevant policies lack depth. Shared policies about weights have more *depth*, if they are themselves grounded in a substantive background rationale or background reasons so that these reasons are collectively accepted.⁶⁷ On the other hand, if the individual reasons for participating in the shared policies diverge, as in the aforementioned case of the company selling vaccines that I have described before, then conflict resolution will be unlikely, since that resolution lacks a common ground.

Before we end this section, I will address an objection raised by Hindriks against the use of Bratman’s shared policies of procedure and about weights to model a group’s normative perspective.⁶⁸ Hindriks claims that joint intentions, such as those discussed by Gilbert (1992) and Tuomela (2013), are more likely to ensure unity and rationality at the group level, than are Bratman’s shared policies, since the former are intentions of the group. I don’t think that this is true. Bratman’s interlocking policies of weights generate consistent and rationally stable beliefs about what ought to be done. Beliefs that the group adopts and realises via intentions. The shared policies of procedures and giving weights are not themselves the mental states of the group. We should rather think of them as constituting aggregation mechanisms that transform the individuals’ inputs and ensure that rational requirements hold for the resulting normative group attitudes and, further, that these attitudes are robustly and systematically generated. So a web of shared policies of giving weight understood via Bratman’s notion of shared intentions seems to be able to ensure the a robust kind of unity and rationality.

⁶⁷Bratman (2013, 145–146)

⁶⁸Hindriks (2018, 10–12)

2.6 Proceduralist Accounts and Supervenience

List and Pettit believe that a group's mental states are determined solely by its members. They present the following argument for this claim.

Given this understanding of organizational structures, how should we expect the performance of a group agent to relate to its members' contributions, that is, their attitudes, actions, or dispositions to act? The things a group agent does are clearly determined by the things its members do; they cannot emerge independently. In particular, no group agent can form intentional attitudes without these being determined, in one way or other, by certain contributions of its members, and no group agent can act without one or more of its members acting.⁶⁹

According to L&P, this statement supports the thesis that the attitudes of group agent's *supervene* on the attitudes of its members: so that there can be no change in a group agent's attitudes without a corresponding change in its members' attitudes. They argue for this claim by restricting their attention to jointly intentional group agents, agents in which aggregation functions can, in theory, be picked out and identified by looking at the organisational structure. That structure, which I have argued above, can be understood via Bratman's notion of shared policies of procedure.⁷⁰ But shared policies of procedures do not capture all the elements that L&P want aggregation functions to capture. Aggregation functions also specify the changes in an individuals' beliefs and preferences that she inputs into the procedure. Shared policies of procedure or an organisational structure only picks out which sub-set of members must decide on a certain task, how they are to decide and which sub-set of members must enact that decision. Aggregation functions are more complex and specify more of the intricacies of the relationship between how the individuals' inputs relate to the group-level attitudes. Importantly, as L&P point out, the output of the aggregation function need not be collectively endorsed in a way that goes beyond what is required to implement the shared policies

⁶⁹List & Pettit (2011, 64)

⁷⁰They also consider the possibility that it is not members' intentional contributions shaped and selected by shared policies of procedure, but that merely mechanical contributions might also give rise to a functionalist system which mirrors a mind, much like the China-body of Block's 1980 thought experiment. As Epstein (2015) shows, this argument rests on the assumption that a mechanical group agent and an intentional one exhaust the possibilities of explaining a group's mind. As I will show in Chapter 3, there is at least one other possibility which takes certain group attitudes to be determined by the legal and social framework in which the group agent finds itself in.

of procedures and attributions of weights. Once we have identified all shared policies of procedure, together with the shared policies of giving weight, we have, in principle, identified all relevant aggregation functions. All aggregation functions, List and Pettit argue, can be understood as a functional relation corresponding to a supervenience relationship.

This metaphysical interpretation of the aggregation function as a supervenience relation, for which the authors do not give an argument above and beyond their understanding of an organisational structure, allows us to characterise these accounts as proceduralist accounts.⁷¹ Proceduralist accounts assume that group attitudes are exhaustively determined by the manifested attitudes of the members, or by the members' attitudes, as is the case with the thicker understanding of group agency by way of group self-identification and the relevant decision procedures that are encoded in the shared policies of procedure and giving weights. Most theories of group agency are proceduralist accounts. They include, among them French (1979), Collins & Lawford-Smith (2016), Collins (2019), Björnsson & Hess (2017) and Tollefsen (2002).

2.7 Summing up

L&P's thin and thick accounts of group agency are a great starting point for analysing and understanding group agency. They clearly depict how, when and under which conditions we attribute mental states to a group and how these mental states robustly satisfy rational requirements. Aggregation functions are central to their account. Aggregation functions build on the rich work in social choice theory and show how certain ways of aggregating decisions – aggregation via a majority function – does not ensure consistent group level attitudes, whereas others, such as a premise-based procedures or functions that are functionally inexplicit like a straw vote procedure, do (recall Section 2.3). Although aggregation functions take up a lot of room in their theory, their analysis is somewhat cursory. In Section 2.4.2, I tried to fill that gap by showing how we can think of aggregation functions as shared policies of procedures. But shared policies of procedure are not all there is to aggregation functions: since L&P also take some aggregation functions to alter the individuals' inputs. Shared deliberation is an instance of an aggregation mechanisms which alters the individuals' input to the aggregation function. Shared deliberation can be understood by means of shared poli-

⁷¹I have adopted the term 'proceduralist account of group agency' from Niels de Haan, who has used it to capture the importance of decision-making procedures in these theories of group agency.

cies concerning weights (Section 2.5.4), which, together with shared policies of procedure, allows us to completely analyse L&P's abstract aggregation functions.

Within this abstract framework of aggregation functions, L&P distinguish their thin and thick accounts by how the group agent counts as reasoning and acting on reasons. The thin account is largely based on aggregation functions that take a broad range of attitudes as inputs, such as normative propositions, and connects them to its members' motivational states in such a way as to prompt beliefs with a normative content that result in the realisation of group actions (Section 2.4.4). Group agents which satisfy the thin account need not have the capacity to see themselves as agents capable of reasoning. They need only display the functional profile associated with normative beliefs.

L & P's thick account posits that, for a group to reason, to self-regulate and understand itself as having these capacities, the members must identify with the group, where such identification reaches the level of group self-identification (Section 2.5.2). I have argued that this is still not yet enough. For robust, diachronic reasoning, self-identification itself only ensures that existing ought beliefs and ends move the agent. They do not, by themselves, guarantee the more sophisticated kind of deliberation and reasoning that is required to reach ought beliefs in the first place. This is ensured by introducing a normative perspective via a web of shared policies of assigning weight to normative considerations which allows and enables collective deliberation (Section 2.5.5).

Both accounts alike take the mental attitudes of group agents to be determined by the members' attitudes (Section 2.6). They provide two proceduralist accounts of group agents, a methodological choice the authors justify by appealing to the role of aggregation functions which are supposed to suffice for exhaustively determining all the mental states which we attribute to group agents.

In the next two chapters, I will present two challenges to both the thin and thick account of group agency. First, neither the thin nor the thick account can distinguish between intentional actions by the group and mere group activities. Second, both the thin account and, to a certain degree, the thick account can ensure that a group person in their sense can be a competent participant in the web of obligations. Thus, a group can satisfy their account, while not being a moral agent.

3 The Attributionist Challenge

3.1 Introduction

One of the advantages of an account of group agency and an account of group action is that it can help us trace the causal chain of events that has issued in an action back to the agent who brought it about. This is central to our practice of responsibility, but also important, when we want to predict future behavior of an agential system. Determining the right locus of agency that has brought something about is particularly challenging in the case of groups, which act through their members. In the case of the group, there are three possibilities of locating agency: The action is due to the group, both the members and the group, or attributable to only a set of individual members. This chapter is interested in the latter two cases. Specifically it is interested in the sort of case, where for some reason, to be explored in the following section, the members' actions can be seen as part of the intentional pattern at the group level, but where it is also true that the group is not in control.

I will proceed as follows. First, I develop a challenge for accounts of group actions and group wills and show why proceduralist accounts cannot provide a satisfying answer (Section 3.2). I then consider possible responses from proceduralists to the objection (Section 3.3) and finally, some of the metaphysical consequences are considered (Section 3.4).

3.2 Misappropriation and External Ends

Before I present my first objection to proceduralist accounts, consider the following ordinary case of action attribution described by Björnsson & Hess (2017, 279-280). It describes how the Acme corporation comes to act intentionally.

Acme's existing commitments to seeking profit, producing industrial chemicals, and being environmentally responsible. Reason-

ing from these commitments, Acme's board members could easily conclude (and vote) that Acme ought to develop a new product – perhaps steel additives. They could also conclude that this should be done in ways that are protective of the environment, but that Acme's commitments to profit and production are more central to Acme and take priority. In this way Acme develops a new commitment – in effect a desire or intention – to produce steel additives, preferably in an environmentally responsible way, and the work goes forward. [...] The members reason from Acme's commitments, not from their own, and Acme's actions thus express its own commitments rather than those of its members. [...] When member behaviour is guided by these corporate commitments (in normal ways), the corporation acts.

Similar to L&P, Björnsson and Hess believe that a group's commitments, goals and other mental states are determined via majority voting or other group decision mechanisms which can be identified by considering the shared policies of procedure and the shared policies of giving weight which have been adopted by the group agent. Björnsson and Hess' account differs from L&P's in that the former explicitly allow for a commitment or a group intention to count as the group's intention, as long as it is generated via the right decision mechanism, without regard for whether it is rational.¹ The picture that emerges seems to be largely plausible. Acme seems to be in control, in that the members' decision to produce a new steel additive is governed and influenced in a normal way by the company's existing goal of making profits by producing industrial chemicals and their additional commitment to doing so in an environmentally friendly way. Consequently, we may, according to List and Pettit's account, describe the production of the new steel additive as Acme's intentional action.

Although L&P's theory and related accounts successfully capture the ordinary actions of corporations, they fail in cases of the following kind.

Misappropriation: A restaurant chain, 'Vienna Schnitzel', listed on the US stock exchange, has three CEOs, Bob, Sue and Ann. Bob and Sue are responsible for the day-to-day management, while Ann is concerned with long-term strategic planning. In their position and role as CEOs

¹While Björnsson & Hess (2017, 279, footnote 17) require that decisions are made from the 'point of view' of the corporation, they do not constrain the elements that constitute this point of view to only those elements that are rationally acquired through deliberation. They also allow for the group's point of view to include irrational ends, thereby circumventing Sugden (2012)'s criticism against L&P. See also Hess (2010, 2012) for similar accounts.

in charge of day-to-day management, Bob and Sue intend to misappropriate the restaurant's funds by arranging to buy the pork used for the restaurant chain's famous schnitzels exclusively from their friend Jane, who charges the firm double the market price in order to secretly transfer some of her profits back to Bob and Sue. Their role as CEOs specifies that they must make these decisions unanimously, which they do, since both benefit from their deal with Jane. Their decision to buy the pork is executed and realized by those in charge of buying the ingredients for the restaurant chain's signature dish.

In the 'Misappropriation' case, the group agent, as a public corporation, is intuitively not in control and, thus, buying the pork from Jane is intuitively not an intentional action of the Vienna Schnitzel chain. Nevertheless, proceduralist accounts must disagree with this statement.² They will have to attribute the self-dealing choice by Sue and Bob on behalf of the corporation to the corporate agent, because Sue and Bob have acted in their capacities and roles as managers in making the purchase. These roles are specified via shared policies of procedure, as well as policies which assign weight to considerations. We can provide an explanation in terms of a corporate decision mechanism that takes as input Sue and Bob's individual preferences to use Vienna Schnitzel to increase their wealth, and translates those preferences into the restaurant's preferences. Therefore, we end up with an intentional profile for the corporate agent, which can be translated into a purposeful action of buying the pork from Jane. Admittedly, the intentional profile will correspond to the story of the corporation acting in a self-destructive way which benefits only Sue and Bob, but it need not thereby violate any formal rational requirements. Indeed, Sue and Bob couldn't care less whether the restaurant suffers: we might imagine that they have 'voted' among themselves to use the restaurant's resources for their own benefit, changing the relevant weight-assigning policies so as to conform with this goal.³ Thus, according to proceduralist accounts, buying the pork is a proper intentional action performed by the group. Call this failure to distinguish between group intentions and group activities the *attributionist challenge*.

One reason for our intuition that the corporation is not in control is aptly captured by the following remarks from Velleman.

²(Hess 2014, 254) appeals to a somewhat similar case as an instance of an external force that undermines a group's agency, but, as I argue in the next footnote, her solution does not apply to Misappropriation.

³Assuming that they have done so gradually by weakening other commitments that might have originally prevented this, would satisfy the condition that Hess (2014, 254) argues for, which require that for a mental state to be the group's own it has to issue from a gradual process from a group's previous rational point of view.

What makes us agents rather than mere subjects of behavior is our perceived capacity to interpose ourselves into the course of events in such a way that the behavioral outcome is traceable directly to us.⁴

Here, the restaurant, as a locus of agential control independent of Sue and Bob's self-interest, is unable to interpose itself because its mind has turned into a mere extension of Sue and Bob's selfish preferences. The corporate 'self' that could interpose itself is crowded out by the individual preference that Sue and Bob force onto the group's mind. The idea underlying this intuition has been endorsed by Harry Frankfurt. He claims that the incapacity to change certain basic commitments, such as the commitment to self-preservation and its higher-order endorsement, are essential to self-governance and, thereby, essential to control.⁵ I will not defend this claim here and only state that this is one possible way to flesh out the idea that, in cases of misappropriation, the corporation is not in control and, thus, not acting in the full sense, since it is not being guided by its own end of self-preservation. I will pick up this strategy, as one way for proceduralists to defend themselves against the objection, later on. The proceduralist's problem is that she has no plausible way to account for the incapacity of changing basic commitments. It is important to note that not every volitionally necessary commitment is required for self-governance or control by the agent. 'Unwelcome ends', due to pathologies such as addiction, compulsion and obsession, can undermine the agency of the agent.⁶

We can see this by considering another argument which explains why misappropriation does not amount to group action. This argument comes from an analogy between the structure of misappropriation cases and a thought experiment about individual human agents gripped by another agent's control. It serves to elicit the intuition that certain mental states which are external to an agent undermine his agency such that we wouldn't attribute a certain action to him and shouldn't hold him responsible for it.⁷ Here is Pereboom's presentation of the case.

⁴Velleman (2009b, 127–128)

⁵Frankfurt & Frankfurt (1999, Essays 9, 11 and 14.). See also Bratman (2018, 105–106) for a discussion of this claim.

⁶See Frankfurt (2006, 43) and chapter 4 for the collective equivalent of an 'unwelcome end' from a moral perspective.

⁷The thought experiment originates from the debate between compatibilists and incompatibilists about free will. It is intended to show that compatibilists are implicitly committed to the claim that Professor Plum is responsible for the killing of Ms. White, even though his actions are guided by an externally induced desire. Although some compatibilists, like McKenna (2014) bite the bullet or 'take it on the chin', accepting counter-intuitively that Plum is responsible for the murder, most, such as Fischer (2004) and Mele

Professor Plum was created by a team of neuroscientists, who can manipulate him directly through radio-like technology, but he is as much like an ordinary human being as is possible, given his history. The scientists ‘locally’ manipulate him to undertake a process of reasoning, directly producing his every state moment by moment, which leads to the killing of Ms. White for egoistic reasons.⁸

We may be inclined to think of the misappropriation case as one in which Sue and Bob pull the strings of the organisation, much like the evil neuroscientists pull on Plum’s strings, so that the corporation is forced, against its own interests, to engage in actions that are harmful to itself. Such an instance of manipulation may be, as Watson writes, ‘intelligibly seen as an impediment [to agency] only if it is seen as in some way external to the self.’⁹ Within the proceduralist framework, the notion that some desires and ends might be external to the corporate self and thereby impede its agency cannot get off the ground, given the proceduralist’s insistence that the group’s point of view is solely constituted by the members’ input. There is no way to ‘smuggle in’ a genuine desire that is not held or manifested by the members who are authorized to form the relevant mental states on behalf of the group. Or, to put it still differently, there is no way to distinguish between the desires which are external to the corporation and those desires which aren’t.

Why should we buy into the proceduralist’s metaphysics? The intentional stance itself is not committed to any such assumption. It only requires the existence of states which play their characteristic functional role. It does not specify where these states are to be located or how exactly they must be realized. Specifically, the stance does not require the states to be manifested by the group members. Surprisingly, L&P seem to agree with this interpretation of the intentional stance. As I showed in Section 1.4, they do not, at the outset, commit themselves to any specific ontology of any agent’s mental states. Only later on do they argue, based on their notion of aggregation functions for judgements, for their supervenience claim.

But where should we look, if not at the members’ contributions? We may find the states at issue by focusing on the social fabric in which a corporation is itself situated, for example, the legal framework, although by itself

(2005), argue that Plum isn’t responsible. Those compatibilists object that the argument, which consists of a series of allegedly analogous cases which terminates at one in which determinism is true, breaks down at some point.

⁸See Pereboom (2006, 112–113).

⁹The idea that desires can be an impediment to agency, if they are external to the individual, is discussed by Watson (1987) and Wolf (1993).

the legal isn't sufficient as I discuss in chapter 5.¹⁰ In the case of US-based corporations, part of this legal framework is comprised by the business judgment rule (BJR). An appeal to the BJR can show why buying the pork is not a full-fledged intentional action by the group by externally fixing the corporation's end. The BJR states the following:

the director shall perform his or her duties in good faith, in a manner he [or she] reasonably believes to be in the best interests of the corporation, and with such care as an ordinarily prudent person in a like position would use under similar circumstances.¹¹

The BJR ensures that when a manager displays sufficient care in choosing on behalf of the corporation in a way that is in the corporation's interest, she is protected from being held individually responsible for the ensuing consequences of the corporation's activity. However, this legal protection is revoked when '[t]he conflicting interest on the part of the deciding directors would taint the transaction and strip the board [of directors] of the protection that would normally be afforded by the business judgment rule.'¹² When protected by the BJR, the board's choices on behalf of the corporation are protected insofar as they count as the corporation's action, such that the consequences which follow from these choices must be borne by the corporation itself. When this protection is withdrawn, the choice may no longer be attributed to the corporation as a collective agent, but is, instead, attributed to the board as a loose group of individuals or to some board member as an individual acting independently of the group. This is effectively captured by Blair and Stout's comment that 'corporate law deems a self-dealing transaction voidable at the firm's option.' The important thing to notice is that the firm does not merely self-correct, like someone making up for a bad deed: the firm can actually 'delete' the action from the set containing all corporate actions, cancelling the thereby accrued commitments.¹³ This assessment captures our previous intuition – about the desire harming the corporate body

¹⁰The basic idea is to appeal to social structural explanations as Haslanger (2016) has called them. For the legal framework to play the role of a structural explanation, it has to be undergirded by relevant social practices.

¹¹See the US Model Business Corporation Act § 35 (1979).

¹²Committee on Corporate Laws, Section of Corporation, Banking and Business Law, American Bar Association, Corporate Director's Guidebook, 33 Bus. Law. 1591, 1603–04 (1978).

¹³To use a term which Brian Epstein coins for actions that do not meet external constraints and, as a result, do not have the powers they would normally have, we might also say that the action misfires. But this seems to presuppose that, in the first place and under slightly different counterfactual circumstances, it actually is an action of the group. This does not seem to be the case if we hold the members' self-dealing motivations fixed.

to benefit Sue and Bob – as external to the corporate agent and as inhibiting or impeding its agency.

A closer look at the wording of the BJR reveals how it does this. The BJR demands that, in choosing on behalf of the corporation from that corporation's first-person perspective, the directors must choose while believing that this choice is also in the best interest of the corporation, or at least not contradicting those interests. Put differently, their choice must be rationalized partially by its coherence with this specific belief about the corporation's best interests. Since self-dealing has, by convention, been taken to conflict with a corporation's best interests. Thus, the interpretation of that specific end by the members of the corporation is not completely flexible. Vienna Schnitzel can't simply by fiat equate the interests of the corporation with Sue and Bob's own selfish interests. Thus, the corporation's self-interest is a legally fixed,¹⁴ external aim which shapes the first-order perspective of any corporation, including Vienna Schnitzel. That self-interest will prompt the corporation to attempt to communicate the nature of the end which prompted Sue and Bob to make their deal with Jane, once Ann and other disinterested members of the board discover what Sue and Bob are up to.

So, the US legal framework (like many other legal systems, for example, the German one) offers tools to distinguish between intentional group actions for which the group, as a separate entity, is responsible from mere group activities for which the members are responsible (perhaps in addition to the group, but not to the same extent as when the group is fully in control and, because of that, fully responsible). It withdraws that legal protection in cases such as the misappropriation case, in which the action isn't being attributed to the group as its autonomous action, but to Sue and Bob.

Given that this is a persistently recurring theme in the social world of group agents, it seems strange that L&P cannot account for it. This way of externally grounding a corporation's ends seems necessary for the group agent to partake in a practice of assigning moral responsibility which revolves around the quality of the will. This is especially true if we care about establishing whether a group's action reflects its will or whether the group has been hijacked by an agency external to it. To do so, we need a principled way of distinguishing between group activities and group actions, on the basis of which the group might be excused for the former and punished more severely for the latter. Proceduralists cannot provide this distinction.¹⁵ L&P, like

¹⁴Although, it being fixed in this way does not only depend on the word of the law, but also on relevant social practices and conventions.

¹⁵Strawson (2008), Wallace (1994), Watson (1996) and Wolf (1987) are just some of the many philosophers who have argued that our notion of responsibility is closely tied with expectations about another person's good or ill will.

other proceduralists including Björnsson & Hess (2017) and Collins (2019), intend their account to show how a group person may be held responsible in situations in which we care about the will of the agents.¹⁶ It is unsurprising that L&P's thin account encounters this difficulty of distinguishing between group activities and full blown actions given that the account tries to capture group actions by a Davidsonian conception of intention. To quote Bratman, that conception affords little room for an understanding of an agent's will "over and above' the operation of event-causal factors'.¹⁷

3.3 The Proceduralist's Response

Proceduralists might take one of two tacks to counter the misappropriation challenge. First, they might claim that misappropriation is, after all, not a proper action on their account and there is some standard of agency which has been violated and which I have failed to mention. Second, they might bite the bullet and maintain that misappropriation is an instance of a full-fledged group action.

There are roughly two ways to argue for the first strategy, which I will discuss in turn. One might argue that, in the misappropriation case, some requirement of rationality is violated, and that requirement is constitutive of exercising one's will and responding to one's reasons as reasons. In other words, one might argue that, in this case, the group is not self-governing because its will does not satisfy constitutive rational standards.¹⁸ Call this strategy the *irrationality response*.

Not all accounts of willing rely solely on the satisfaction of rational requirements to identify whether some group behaviour is a will or a mere activity. Hierarchical accounts of action such as Pettit's account of group self-identification introduce emotional, affective features – like investing into a group identity, feeling shame or pride on behalf of the group – as being constitutive of endorsing or owning a group action. These more substantive accounts might argue that some feature of misappropriation cases undermines group self-identification on the part of those individuals acting and deciding on behalf of the group, which means the group's actions are not fully intentional. Call this strategy the *substantive response*.

Irrationality Response: One rational requirement that might be violated

¹⁶See Section 4.2.5 for a more detailed argument why L&P are committed to such an understanding of responsibility.

¹⁷See Section 2.4.4.

¹⁸See Velleman (2009*b*, Introduction) for an argument to the effect that this is the mark of a fully intentional and autonomous action, but also Bratman (1999, Chapter 10).

appeals to the irrationality of giving up certain ends. Specifically, proceduralists might claim that it is irrational for Sue and Bob to adopt the aim of turning Vienna Schnitzel into a self-destructive money making machine. This choice is irrational because, in adopting it, the aim of corporate survival is thereby dropped, contrary to the demands of rationality. This makes the subsequent action of buying the meat at above-market price similarly irrational. The irrationality of abandoning the aim of corporate well-being is due to the fact that this aim is a necessary or constitutive element of the point of view of the corporation.

There are two possible sources for this necessity which the proceduralist may invoke: one conceptual, the other practical.

- (a) The supposed conceptual necessity follows from the claim that it is part of the very concept of a corporation that among its aims is the aim for its own survival. But why should this be so? A state-owned corporation which produces some public goods might, by the very nature of those goods, not be profitable or be led in an unprofitable way, being kept alive through government support. Furthermore, there are lively discussion about the ends of governance in the business ethics literature: there is nothing close to an agreement about the ends of a public corporation, at least academically. Alternatively, one may argue that the end of survival is a constitutive part of the notion of a corporation because the corporation's existence is intrinsically valuable. But it is clear that corporations can create many bad outcomes, and that any corporation's supposed value is strongly contingent on what it produces and how. Thus, the value of a corporation is, unlike the dignity of persons, derivative and not a matter of conceptual necessity.
- (b) There may also be a second, practical source of the necessity of the aim of corporate well-being. Because many people have invested their energy and resources into the corporation, they are, as a whole, committed to the end of keeping it alive. But, as L&P have rightly pointed out, the outcome of an aggregation mechanism need not be accepted by every member of the group agent whenever the mechanism plays its functional roles and leads to action – as, for example, in the misappropriation case. We might also consider the possibility that the group has previously, perhaps at its founding, chosen to commit itself to the end of furthering

its self-interest in its future actions. But this, too, is not enough, since any commitment can also be discarded or interpreted in a way that rationalizes the self-dealing end. If these commitments have tied the group up in commitments towards non-group agents, which they cannot give up unilaterally, the source of these commitments is again outside the agent. Furthermore, a commitment need not translate into an action, but, rather, plays the role of an element in the corporate agent's deliberation. This leaves a third, externally fixed, source for this necessity. As I have argued before, because L &P and other proceduralists are committed to the supervenience of group attitudes on individual attitudes, this option is unavailable to them.

The Substantive Response: There is a further response open to proceduralists who, like Pettit, defend a thick account of group agency. They might argue that, in *Misappropriation*, buying the pork is not an instance of the corporation's will, since Sue and Bob do not identify with the point of view of the corporation. Even if they do so identify in some limited sense, they are clearly not invested in the corporation in that they do not take the corporation's ends as making 'the same intimate claims' on them or on their individual identities. The problem with this line of argument is that, on its own, group self-identification doesn't restrict Sue and Bob from including into the corporation's point of view a self-destructive motivational element. We might even think that, although they are invested in the corporation insofar as its previous success makes them, the corporate managers, look good and powerful, they are even more invested in their own personal identity – which may consist in having a big villa and buying a bigger yacht than their neighbours. Perhaps it may be argued that the end of self-destruction is incompatible given the history of the restaurant's management. Previous managers might have, in a Protestant manner, given their all and sacrificed their individual identities in favour of time and effort put into their work for Vienna Schnitzel. This history would only put some substantial pressure on Bob and Sue if it were brought up by someone superior to them within the hierarchy. But since Bob and Sue are ultimately responsible for the Vienna Schnitzel company's daily business decisions, this pressure does not affect them. Even if their actions alienate many group members at lower levels of the hierarchy, that alienation need not impact the status of the self-identification with the relevant choice to misappropriate, which, if rationalised, implies the corporation's demise. A more explicit account of the emotions involved

in investing in a group identity might serve the thick account with additional resources to defend against instances of misappropriation. But I suspect that this will be difficult, since emotions do not come with substantial or rational requirements which restrain their application.

With these possibilities exhausted, proceduralists might take the other, bullet-biting tack mentioned earlier. This approach is possibly endorsed by Björnsson & Hess (2017), who accept that the point of view of a corporation can contain irrational elements. So, proceduralists might own the consequence of this argument, and accept that misappropriation cases indeed describe corporate actions. They will argue that the BJR is a mere contingent legal fact which should not be part of a theory of action. The social status of a group action, legally rescinded in this case, alone indicates nothing of the action's metaphysical status. How might one argue for this claim? Proceduralists might point at a seemingly analogous individual case. Suppose that I intentionally sign a contract which includes a clause by which I enslave myself to my contractual partner. This clause renders the contract moot, so I am unsuccessful in signing the contract. Although my action misfired and failed to have the relevant social status, I nevertheless intended to sign the contract. In this case, we might claim that I at least signed the contract, although my action did not achieve the social status of a successful contract signing. But the point at issue in misappropriation isn't the absence of the right social status, but the reasons for this absence. The analogy gets wrong the reason why my action was not assigned the expected social status. If I were not in control of myself at the time of contract signing – because I was being controlled by another entity or not in control of my cognitive faculties – then the contract case would be a valid analogy. But I am in control: perhaps I even wanted to enslave myself. In the misappropriation case, the action's social status is much more entwined with the status of the group's agency than it is in the individual case. The group agent is, after all, itself an essentially social entity.

3.4 The Misappropriation Objection and its Metaphysical Consequences

If the misappropriation objection succeeds, it challenges the metaphysical assumption that group attitudes supervene on individual attitudes. This puts the objection in the dialectical neighbourhood of an argument raised by Brian Epstein in *The Ant Trap*. Epstein's challenge is developed within a metaphysical framework for social ontology inspired by the physical sciences.

To discern the existence and form of social kinds – things like money, marriage, groups or group actions – he introduces two metaphysical relations: grounding and anchoring.

The grounding relation says something about how certain social kinds or facts about constructs like money are metaphysically explained or ‘made’ by other, more fundamental social kinds. For example, ‘the fact “The barn door, walls, and roof are burning” grounds the fact ‘The whole barn is burning’; the fact ‘I am married’ grounds the fact ‘I am not a bachelor’; and the fact ‘a million herring turned in such-and-such directions’ grounds the fact ‘the school split in two’.”¹⁹ Analogously for social kinds: for example, the fact that a bill X was printed by the Bureau of Printing and Engraving grounds the fact that X is a dollar bill.

Epstein thinks that, in all these cases, the relationship ‘facts X ground facts Y’ expresses the logical property of sufficiency rather than necessity or determination. We can discover whether some sets of facts are grounded by other sets of facts by looking at whether the first set co-varies with the second. The supervenient relationship between group attitudes and the attitudes of individual members also entails that group-level attitudes are grounded by individual member attitudes. The evidence for this claim is taken to be the fact that a change in the group’s attitudes entails a change in its members’ attitudes towards the group.²⁰

Frame principles govern how certain *kinds* of facts ground other types of facts. They are generalisations or rules about the kinds of facts that ground other kinds of fact.²¹ When we look at a social phenomenon such as a group action, we might think that there are many different possible frame principles which may explain what a group action is and how it is connected to more fundamental notions. A proceduralist will identify a different frame principle as the grounding relation than I will propose in Chapter 5 about my constitutive account of group agency. Epstein calls the question about which frames and grounding relations obtain ‘the anchoring inquiry’. Anchoring puts the frame principles in place: it reveals why some set of rules, rather than other facts, explains some social entity. There are different kinds of facts that anchor social entities. Sometimes, collective acceptance – which may be understood as some form of shared intention discussed in Section 2.4.1 – anchors a certain frame. Sometimes, a social fact is anchored by a mere aggregation of individual properties, such as the fact that millenials are

¹⁹Epstein (2015, 70)

²⁰See also Epstein (2015, 72) for other kinds of evidence for grounding relations.

²¹This kind of universality is displayed in the following quote from Epstein (2015, 78). ‘A frame principle,’ Epstein writes, ‘gives the grounding conditions not just for the actual world, but for all possibilities.’

likely to work non-permanent jobs.

Within this framework of anchors and grounds, Epstein discusses the following case. It displays certain parallels with mine.

Among the powers anchored for the chief justice is the power to assign authors for opinions of the court. Suppose Justice Scalia gets too big for his britches, and sends out an email to his colleagues assigning responsibility for writing a case's opinion to Justice Thomas. Suppose that Chief Justice Roberts sends out an email assigning responsibility for the same case to Justice Kagan. Then suppose both Kagan and Thomas publish opinions. Kagan's publication is the issuance of an opinion by the court. Thomas's publication is not. It is not enough to fix the membership of the Supreme Court and what those members did, in order to determine the action of the court. Even granted this, we need to ground the further fact that Roberts is chief justice. External facts—not just facts about members—ground the fact that Roberts has differential power. The role of chief justice is anchored by the constitution and Judiciary Acts. But the fact that Roberts fills that role has external grounds. These external grounds are among the facts that ground the issuance of opinions by the Supreme Court.²²

The parallel between Misappropriation and the Supreme Court case is clear. In both cases, the legal framework, being external to the group itself, can help us identify what counts as an action performed by the group. The Misappropriation example is helpful insofar as it can best be understood by introducing an externally fixed end, which makes a difference not only for who makes a decision, but substantially influences the outcome of decision-making. However, the most important takeaway is that both cases indicate that a theory of group agency and group action must – or at least may – contain some elements external to the group itself. The 'must' at issue is of a normative kind: in order for a group to avoid becoming a puppet for the benefit of a subset of its members, it needs a practical core that resists unilateral change. The 'may' highlights that, as a matter of fact, certain groups are externally constrained. Therefore, if we want to explain how these groups function and distinguish, in practice, between group actions and activities, we must abandon the supervenience relationship in favour of Epstein's more permissive, less restrictive ontological framework.

²²See Epstein (2015, 223).

3.5 Summing up

I have argued that proceduralist accounts of group agency cannot distinguish between group actions, which can be attributed fully the group agent as an independent locus of control and mere group activities. Cases that bring this to light are characterised by rogue members that act for their own interests in a way that undermines the group's interests. I hope to have made plausible that in our social world we answer the question of a group's own interests by reference to facts outside the group agent, such as the social and legal practices in which the group is embedded. If this turns out to be the right way to distinguish between group actions and activities, then proceduralist accounts turn out to be ill suited for providing an account of a group will.

4 Group Responsibility and the Pathological Group Agent

4.1 Introduction

L&P belong to a camp which takes an agential system's capacity to be held responsible as a requirement for personhood. As I have said before, this means that their account of group agency is actually an account of group personhood. So far, we have taken for granted that their thin and thick accounts provide us with a model of an agent as a competent participant in the realm of obligations. But we have not yet considered the practical dimension of responsibility, in which these obligations become operative as demands placed on the group agent and as the content of claims that a group agent might direct at itself or other agents. These demands and claims, along with the moral expectations that correspond to them, are embedded in and shaped by the practice of responsibility.

This chapter will explore and settle on two ways to understand and articulate the practice of responsibility within a broadly Strawsonian picture of responsibility as constituted by the reactive attitudes (Section 4.2 - 4.2.4) apply it to L&P's conception of group responsibility (Section 4.2.5 and consider a challenge for their account, which highlights the fact that their understanding of group agency does not account for characteristic structural defects in groups (Section 4.3 - 4.4.1).

4.2 The Reactive Attitudes and the Quality of the Will

In his landmark essay 'Freedom and Resentment', Strawson argues that understanding moral responsibility requires understanding its place and role in ordinary interpersonal relationships. Interpersonal relationships are shaped by a number of different emotions, including shame, anger, indignation, resentment, guilt, hate, disgust, appraisal, gratitude and forgiveness. Strawson

notes that these emotions are natural responses to the violation or satisfaction of certain kinds of moral expectations. They show ‘how much we actually mind, how much it matters to us, whether the actions of other people – and particularly some other people – reflect attitudes towards us of good will, affection, or esteem on the one hand or contempt, indifference, or malevolence on the other.’¹ Strawson calls these *participant reactive attitudes* because 1) they are natural *reactions* to others’ good, ill or indifferent will and 2) they are integral to being in an interpersonal relationship or regarding another person as a potential partner for such a relationship.² Therefore, Strawson invites us to understand responsibility in terms of holding others responsible and in terms of the role and function of participant reactive attitudes.

This function may be variously observed. For example, guilt is a reaction to an agent’s own act or omission, which elicits anger, resentment and indignation on the part of those agents affected by its consequence. Then, the feeling of guilt guides the agent towards those who have been harmed and motivates her to offer reparations or inflict some punishment on herself.³

In general, the reactive emotions also come into play when we consider different aspects of our practice of holding each other responsible.

If someone treads on my hand accidentally, while trying to help me, the pain may be no less acute than if he treads on it in contemptuous disregard of my existence or with a malevolent wish to injure me. But I shall generally feel in the second case a kind and degree of resentment that I shall not feel in the first.⁴

Other, more subtle cases are easy to come by. If I expect you to buy me flowers for my birthday, but must convince or persuade you to do so, I will not feel the same way receiving them as I would had you done it without needing to be prompted.⁵

These cases highlight how reactive attitudes mirror or correspond to degrees of goodwill, or the absence of it. But they also explain paradigmatic cases of excuses and exemption, such as when I accidentally step on your hand.

¹Strawson (2008, 342)

²The *Stanford Encyclopedia of Philosophy* article by Eshleman (2016) provides a handy guide to the central features of the Strawsonian account.

³See Williams (2008, 89–90) and Williams (2008, 94).

⁴(Shafer-Landau 2012, 342)

⁵Shoemaker (2011) uses this example to illustrate the different kinds of moral or normative standards that interpersonal relationships consist of.

4.2.1 Excuses

Why do these cases justify modifying or suspending our reactive attitudes? Typically, this is due to the availability of excuses. According to Austin (1971), in excusing herself for doing X, an agent aims to show that, even though, under the circumstances, X was wrong, it was not done intentionally. If we understand excuses this way, our hand-treader is excused. Other types of excuses are inadvertence, mistake, coercion, necessity and duress.⁶ But why do excuses justify suspending or modifying one's reactive emotions? The rough idea is that, in not acting intentionally, the person who harms us does not choose to act in that way. She does not act on the basis of reasons and, thus, her bodily movements are not sufficiently controlled or motivated by reasons. If we take acting on a good or a bad will to mean acting intentionally for reasons which one takes to be motivating and which one endorses, then excusing conditions capture the sense that an action or behaviour that was neither motivated nor endorsed by the agent does not manifest a bad will. Thus, by highlighting the important role of the quality of the will in assigning responsibility, the Strawsonian framework effectively captures why people typically withdraw their reactive emotion, or have some reason to do so, when presented with an excuse.

4.2.2 Exemptions and the Objective Stance

Excuses present cases in which we temporarily suspend our participant reactive attitudes without suspending our reactive attitudes towards the agent in general. In another range of cases, we never adopt an interpersonal stance in the first place. When interacting with children, the mentally insane or the elderly, we adopt, instead of a stance of involvement, an objective attitude. Strawson writes:

To adopt the objective attitude to another human being is to see him, perhaps, as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained; perhaps simply to be avoided.⁷

The underlying idea is that a lack of certain capacities may render an agent unfit, unsuitable or incapacitated as a potential partner for the kinds of relationships which are characterised by participant reactive attitudes. Here,

⁶See Wallace (1994, 136ff) for a detailed analysis of these different excuses.

⁷Shafer-Landau (2012, 344)

some difficulties arise. In the case of exempting conditions, a question which I have yet to address becomes acute: What kind of relationship do the reactive attitudes point at exactly? Do the reactive attitudes shape our *ordinary* human relationships, as Strawson seems to suggest at some points in his article, an idea that Bratman (1999) has developed further or do they govern a kind of more normatively demanding relationship, namely our *moral* relationships? The latter interpretation seems to be suggested by the following quote.

If your attitude towards someone is wholly objective, then though you may fight him, you cannot quarrel with him, and though you may talk to him, even negotiate with him, you cannot reason with him. You can at most pretend to quarrel, or to reason, with him.⁸

According to Strawson, what is really at issue in those cases in which we exempt someone from the reactive attitudes is their inability to participate in moral communication, co-deliberation as well as the giving and accepting of criticism and justification. Interpretations like this one have been adopted by a number of philosophers including Scanlon et al. (1998), Wallace (1994), Wolf (1987) and Watson (1996).⁹ This interpretation is attractive for analysing group agents, because we are less interested in whether we can have ordinary relationships with them than we are in whether they can competently participate in the realm of obligations and whether we can hold them accountable for their behaviour as we hold responsible individual moral agents. An affirmative answer to that question would allow us to extend our ordinary practice of responsibility to them. But even if we do find that certain types of groups are incapable of participating in moral relationships in which the will and moral communication are essential, there might still be a way in which parts of our practices are relevant to their behaviour and can inform our interactions with them.

That part of our responsibility-assigning practice is constituted by reactive emotions like shame, hate, disgust, contempt and derision. They can result in avoidance and other sanctioning behavior. Although these reactive attitudes might be caused by an agent's act or omission, they need not be. Bernard Williams emphasises this in his fascinating analysis of the ancient Greek understanding of responsibility and agency in his *Shame and Nec-*

⁸Shafer-Landau (2012, 344)

⁹This interpretation is also supported by the difficulty of identifying what an ordinary, interpersonal relationship is. After all, relationships with our therapists seem to straddle the boundary between the objective and the interpersonal stance.

sity.¹⁰

What arouses shame may be some failing or defect [in character]. It will lower the agent's self-respect and diminish him in his own eyes. [...] More positively, shame may be expressed in attempts to reconstruct or improve oneself.

One thing that a marked contrast between shame and guilt may express is the idea that it is important to distinguish between 'moral' and 'nonmoral' qualities. Shame itself is neutral on that distinction: we, like the Greeks, can be as mortified or disgraced by a failure in prowess or cunning as by a failure of generosity or loyalty.¹¹

This second set of reactive attitudes does not track the quality of the will, but properties that the agent has no control over. Nevertheless, they play an important role in our normal practices of responsibility. They are, as David Shoemaker points out, concerned mostly with attributability. The concept of attributability effectively captures the idea that certain actions do not express one's will, but speak for the agent in some sense despite being out of the agent's hands. They do so by expressing those practical commitments which are shaped by the agent's character and emotions.

Our responsibility practices include the relevant sorts of moral responses to mere attributability. Specifically, I have in mind aretaic appraisals, judgments about the morally relevant aspects of an agent's character in light of the agent's attitudes or actions. Aretaic appraisals are familiar. We may judge of someone, in light of some action or attitude, that he is 'cowardly,' 'generous,' 'cruel,' or a 'hopeless romantic.' What occurs here is evaluative, agential predication. [...] This is not mere superficial assessment (as in the deployment of predicates like 'tall,' 'skinny,' and 'pretty') for it goes to the heart of one's actual motivations qua self-expressing agent, implicating the set of psychological elements that provide an explanation of one's motives and attitudes in normative domains. These are the elements with which one is identified, whether or not one is conscious of, or endorses, the identification in question.¹²

¹⁰Greenspan (2003, 2016) suggests that these emotions are appropriate in cases where we judge an agent's character.

¹¹Williams (2008, 94 and 91–92).

¹²Shoemaker (1996, 612–613)

Aretaic judgements can overlap with our judgements regarding another agent's will. Character traits which influence our moral agency are being cowardly or courageous, opportunistic or principled, alert or absent-minded, self- or other-centred, showing constraint against acting impulsively, being cruel or compassionate. These traits influence how well one can satisfy the moral demands imposed on one, they play a role in determining one's moral character or identity, they are relevant to our ability to participate in a moral or ordinary relationships. Thus, they are relevant for one's moral assessment and, in turn, the reactive attitudes. In so doing, they also play a further role in our interpersonal interactions.

[Aretaic appraisal] is a sort of appraisal that still has the sort of wide ranging practical payoff that is typically the mark of responsibility assessments, for we make all sorts of decisions about who we will interact with, or spend any significant amount of time with, or work on loving more, on the basis of such assessments.¹³

So, the reactive emotions do not cover moral relationships alone: they also cover the non-moral aspects of these relationships and of the related agents. Thereby, they also shape those relationships which do not impose on their participants those stricter requirements which are pre-requisites for moral arguments, criticism and justification.

In what follows, I will look more closely at the exempting conditions for moral relationships characterised by the reactive attitudes. Then, I will present L&P's understanding of responsibility, which they claim is in the spirit of the Strawsonian framework. I argue that this understanding falls short in important aspects. To show this, I construct a type of group agent which meets L&P's condition for responsibility and for being a group person in the thick and thin sense, as outlined in the previous chapter. However, this group agent fails to meet certain conditions which some of the main proponents of the Strawsonian framework take to be necessary for participating in moral relationships.

4.2.3 Moral and non-Moral Relationships

What is required to participate in a moral relationship? In moral relationships, we form expectations about obligations that must be action guiding for the relationship participants. They must be able to self-motivate and, when others are held to obligations, these obligations must, in principle, be motivating for the other person. This means that a moral agent must be able

¹³Shoemaker (1996, 615)

to grasp, understand and be moved by the moral reasons that make up the obligations that she is held to. Wallace (1994), in an influential Strawsonian account developed in his book *Responsibility and the moral sentiments*, has referred to these capacities as the powers of ‘reflective self control’.¹⁴

Definition 4.2.1. *The powers of reflective self-control*

Reason-Responsiveness: *Being responsive to moral reasons such that they are able to govern and guide one’s thought and action in the right way until, ceteris paribus, the action is fulfilled.*

Reason-Comprehension: *Certain cognitive and affective resources are needed to recognise and understand morally salient aspects like pain and suffering via the ability to empathise with others, or cognitive capacities to understand more complex moral concepts such as honesty and promising.*

Reason responsiveness is a ‘capacity for critical reflection’ which refers to the ability to assess one’s immediate desires in light of the moral reasons one accepts.¹⁵ This entails the capacity to *control one’s impulses* and, thus, the ability to avoid acting simply on those desires which are the strongest at any given moment. It also involves the ability to *make choices* based on deliberating about the moral reasons one believes to apply to a particular situation. Finally, once a decision is made on the basis of the reasons one accepts, for the agent’s actions to express her choice, she must be able to translate that choice into action, so her intentions must be capable of *guiding her thoughts and bodily movements* in such a way that instantiates her choice. Importantly, this can often entail diachronically extended activities which the agent must coordinate over time.

Comprehending reasons requires that one is able to apply principles which ground moral reasons in a variety of different situations and anticipate the demands which those principles entail in the various circumstances to which they apply. But correctly applying moral principles such as truthfulness, non-maleficence or promise-keeping to a wide range of situations demands a sophisticated understanding of the moral concepts which feature in them, like harm, happiness, truthfulness, commitment and others. In other words, it requires that one knows how these principles could be justified and how actions relate to them.

It is clear why these two conditions are needed to engage in genuine moral discourse. Reason comprehension is important to justify one’s actions when

¹⁴Wallace (1994, 157ff)

¹⁵Wallace (1994, 158)

these actions have fallen short of expectations and demands. That justification can only be carried out by appealing to the relevant moral norms, the morally salient aspects of the situation in which the decision was made and resolving moral conflicts in a reasonable way. This is often referred to as the *answerability* aspect of holding agents responsible. Reason responsiveness, on the other hand, is relevant so that criticism and realisations drawn from moral discourse – maybe regarding reparations which are owed to wronged parties – do not stay at an epiphenomenal level, but are generally capable of moving the agent directly.

According to Gary Watson, much less is needed for aretaic assessments.

Aretaic appraisal clearly requires some normative competence on the part of the subjects of that appraisal. The subjects of moral appraisal in particular must have the capacity to apply moral judgements to their deliberation. Aretaic appraisal requires the intelligence and sensibility to comprehend at least the normative concept in terms of which the relevant forms of appraisal are conceived. (For example, one is open to the appraisal as cruel only if one has the concept of cruelty.) There is no reason to think that this competence must include the ‘ability to acquire the right values’.¹⁶

Following Watson, I will argue that in our practice of responsibility, aretaic judgements with morally relevant content are only justified if the agent can grasp the concept used in an aretaic judgement.¹⁷ This allows us to form aretaic judgements about an agent who acted badly even if the agent didn’t have the powers of reflective self-control.

4.2.4 Two Ways of Holding Responsible

We can now identify two distinct, albeit overlapping ways of holding agents responsible. Following Watson and Shoemaker,¹⁸ when an agent has the

¹⁶Watson (1996, 230)

¹⁷Thus, for example, cases in which we describe a lion as beautiful or cruel are no counterexample to Watson’s claim, since lions are not the target of the reactive attitudes.

¹⁸See Shoemaker (2011) and Watson (1996). Although Shoemaker’s approach to the different aspects of holding responsible is more fine grained, distinguishing *answerability* – those commitments that move the agent and that the agent takes as reason-giving – from *attributability* – which refers to the whole range of, partially emotional and irrational commitments that prompt the agent to act – and from *accountability* – the kind of responsibility agents can only fulfil if they have some version of the power of reflective self-control.

powers of reflective self-control and there are no excusing conditions present, it is answerable and we hold him or her *accountability-responsible*. In that situation, it is fair and appropriate to blame and sanction the agent.

Conversely, when these powers are somehow absent or diminished in an agent, we do not regard the agent as a full participant in the realm of obligations, or as a full member in moral relationships in which one exchanges moral arguments and justifications for one's actions. This is nicely captured by Watson in the following quote.

Seeing the criminal as itself a victim will not prevent us from shutting the cage or pulling the trigger. But these responses will then seem regulative rather than retributive. In a disconcerting way they will lose their expressive function.¹⁹

Whereas in agents possessed of the powers of self-control, the manifested reactive emotions have a communicative role in cases where obligations are violated, this role lacks its intended functions when the agent either does not comprehend reasons or has some motivational deficiency. In these cases, the regulative role of the reactive emotions is to control and manage the agent who is exempt from moral relationships. It is unfair, in these cases, to demand that they manifest a good rather than an ill will, since they are incapable of a good will. They are not the appropriate target of those reactive attitudes in that the attitudes cannot play their usual role. Nevertheless, there will still be reasons to target agents like psychopaths with reactive emotions concerning their character: doing so might supply them with self-interested reasons to avoid situations in which the attitudes of others make it impossible for them to stand in relationships with them. In these cases, we can still form *aretaic-responsibility* judgements regarding an agent's cruel or irascible nature, fickle character or egotistical ways. Character traits that have led them to perform an action for which we now show hate, derision or contempt and which can undermine our social ties with the person. Thus, aretaic-responsibility involves the *attribution* of an action to an agent and a corresponding value judgement regarding the ends or the character that gave rise to it. The object of this judgment will take our reactive attitudes as evidence that they have violated an obligation, even if they will not be motivated by the obligation which they were under. Only the knowledge that others care about these obligations might supply them with indirect reasons for actions.

Let us now situate L&P's working theory of responsibility within this Strawsonian framework.

¹⁹Watson (1996, 230)

4.2.5 List and Pettit's Account of Group Responsibility

When presenting their theory of responsibility, L&P seem to have a Strawsonian account in mind. They write that 'blaming involves adopting or identifying with the stance of a creditor: someone to whom at least an apology is owed. Adopting such a stance typically means indulging in resentment, identifying with the stance means indulging in indignation. Adopting the stance towards oneself, as in blaming oneself, means indulging in a sense of guilt.'²⁰

As we have seen in Section 4.2.3, these are the core reactive attitudes which point at the accountability interpretation of holding responsible. The underlying relationship is a moral relationship within a community of moral agents in which the quality of the will and corresponding answerability demands and moral communication play a central role. This is also supported by L & P's claim that personhood requires being a competent member in the realm of obligations.

For the careful reader, this might come as a surprise: I have argued previously that, at least under their theory of thin group personhood, L&P do not provide an account of a group will in a robust sense which goes beyond the operation of event-causal factors. In Section 2.4.4, I argued that their Davidsonian account of group actions that derives from their thin account of agency does not give rise to an account of a group will having a volitional and, more importantly, a reasoning-centered aspect. Similarly, their thick account of group personhood which includes group self-identification does not ensure this volitional and reasoning-centered aspect of the will without also assuming a normative group perspective by introducing a structure that allows for group deliberation, as we have seen in Section 2.5.5. Furthermore, their account, like other proceduralist accounts, fails to distinguish activities like the misappropriation case discussed in Chapter 3, in which the group is not in control, from intentional actions where the group is in control. Thus, their account cannot help us track the quality of a group's will, because it can't identify a group will in the first place.

But there is a further problem for their account of responsibility. To see this, consider their conditions for the fitness (appropriateness) of holding someone responsible for a choice.²¹

Definition 4.2.2. *A group agent G is fit (in List and Pettit's sense) to be*

²⁰See List & Pettit (2011, 154). Using the metaphor of a creditor recalls the account of responsibility developed by Wolf (1993). In general, most accounts of group agency build on a Strawsonian understanding of responsibility. Examples include Tollefsen (2002), Björnsson & Hess (2017), Hindriks (2018) and Szanto (2016).

²¹See List & Pettit (2011, 155–156).

held responsible for a choice iff

Normative significance: *G faces a normatively significant choice, involving the possibility of doing something good or bad, right or wrong.*

Judgemental capacity: *G has the understanding and access to evidence required for making normative judgements about the options.*

Relevant control: *G has the control required for choosing between the options.*

L&P take these conditions to be individually necessary and jointly sufficient for holding someone responsible. On the face of it, each one is plausible. According to the authors, an agent is not fully responsible if the normative significance condition is not satisfied, that is, if the agent did not face a choice between right or wrong.²²

If the judgemental capacity condition is not satisfied – perhaps because the agent was unavoidably ignorant of the relevant facts – then we would typically excuse the agent.

Finally, we would only hold someone responsible for those choices that are under their control. Control might be understood in the thin sense of causal control over the content of one's normative judgement, or in the thick sense of the normative judgement playing an appropriate role in guiding the agent's action in those cases when group members self-identify with the group.

It is not a straight forward matter that L&P's own account meets these three conditions. For example, the thin account demands only that a group is capable of forming normative judgements, but, as outlined in Section 2.5.4, these judgements can be formed in an *ad hoc* manner. This can only be countered by assuming a normative perspective both for the thin and thick accounts of group agency. This means that L&P's own account would not meet the judgemental capacity without the addition of a normative perspective, a structure which enables shared deliberation via shared policies of giving weights to normative considerations. So, suppose we add this structure to L&P's theory of group agency. Can we expect the resulting system to produce a group agent that satisfies the three conditions of L&P's theory of responsibility?

²²By keeping the condition of normative significance vague, L&P do not commit to any particular moral theory. Therefore, the first condition does not provide any normative recommendation to an agent regarding which specific choice is good or bad. Nor does the condition restrict the kind of moral theory which could be applied. In the individual case, this could be an egoistic theory, while in the corporate case it could be a consequentialist theory which takes the amount and distribution of profits as the salient 'good' to be maximised.

We can, but this does not mean that the resulting group agent would also meet the conditions of reflective self-control. Specifically, we can construct a group agent with a normative perspective that is capable of deliberating and meeting L&P's conditions of agency and meeting the responsibility conditions in definition 4.2.2, given a plausible interpretation of the control condition. At the same time, that agent might fail to meet the conditions of reflective self-control. Such an agent would be responsible, given L&P's definition of its actions, but exempted from the moral community understood in the standard Strawsonian way as outlined in Section 4.2.3.

I will construct such a group agent in the next section. My challenge bears some similarity to an argument developed by Hindriks (2018). There, Hindriks argues that L&P provide an account of a rational group agent that, because it is itself incapable of *moral* deliberation and argument, does not meet the conditions for genuine moral agency. His example alludes to the idea that certain groups are not in control of their ends or that they might have ends that are somehow irresistible to change and revision. Unlike me, Hindriks argues that this can be solved by requiring that the members of group agents with a normative perspective support that perspective with relevant shared emotions such as collective guilt, indignation and resentment. I will argue that the true problem lies with the kind of ends that certain group agents have. In the next section, I construct a group agent beginning with the intuition that some group agents face a particular kind of constitutive, unalterable end which human beings don't encounter and that undermines their capacity to participate in Strawsonian moral relationships.

4.3 Second Challenge: The One-Dimensional Valuer

The challenge proceeds from the following intuitive notion.

One possibility is that group agents, unlike individuals, are frequently, perhaps even typically, created to achieve specific ends. They are, in that respect, instrumental for individual agents, in a way that individual agents are not instrumental for any others.²³

Corporations are brought into existence with the aim of selling a product for profit. NGOs are brought into existence when people want to solve a specific social or environmental problem. States are formed with the aim of establishing a just and authoritative system of laws to solve collective action

²³Chiao (2014, 14)

problems and protect the basic rights of citizens. Sports clubs are formed to provide a space for members and non-members to pursue sports-related activities. Only very few groups – the assemblies of friends or lovers – might be formed just for their own sake or for their own intrinsic value.²⁴

Most theories of group agency assign no particular normative status to the ends that are the reasons for groups being brought into existence, and that are shaped and partially dictated by the legal system and social conventions they exist in. This contrasts starkly with the approach taken in applied ethics, including business ethics, where theorists go to great lengths to consider how the function of a for-profit company influences other potential obligations it is under.²⁵

I will argue that the idea that groups have these functions, which I will describe as ends that are *constitutive* for them as the agents that they are, can pose a unique challenge. This challenge is captured in the following two claims.

- 1) Constitutive non-moral maximising ends exempt groups that are valuing agents from the membership in moral relationships understood via the powers of reflective self-control and
- 2) nonetheless, these groups can be the target of aretaic appraisals.

As I lay out the argument, I will have to modify both of these hypotheses. First, groups with constitutive non-moral maximising ends are not automatically exempted from responsibility. Rather, under certain conditions, they will be under rational pressure to turn into agents that do not meet the conditions required for participation in moral relationships. More precisely, groups with these aims are under rational pressure to be a certain kind of valuing agent – namely, an agent that always prioritises its constitutive, non-moral goals. I call these agents *one-dimensional valuers* (ODVs). In Chapter 6, I will return to the question of whether there are real-world group agents

²⁴In the case of a group of friends formed without any aim, it might prove impossible to adopt the intentional stance. Instead, we might better understand the group as simply acting together jointly, albeit for reasons derived from their friendship. Rovane (1997) discusses the possibility of a couple who synchronise all of their mental states so as to achieve maximum agential unity. If, in such a constellation, all actions and communicative acts are guided by their collective point of view, they, too, count as a group agent in the sense that I have adopted in this dissertation. Given that no two individuals feel and think the same way about every possible decision, such a couple will also need decision mechanisms or aggregation functions which break the tie in the face of conflict.

²⁵Friedman (2007) is the most famous example of the many theorists who try to square the function of firms with possible obligations that they may be under. I will discuss the arguments that circulate in these debates at length in Chapter 6.

with a functional structure and normative profile of an ODV. For now, I will consider them as a mere theoretical possibility.

4.3.1 Constitutive Ends and the ODV

In this section, I will define the one-dimensional valuer. To do this, I must first define a constitutive aim.

Definition 4.3.1. *E is a constitutive aim of an agent G of type φ iff (1) in acting, G aims constitutively at E and (2) if G were to systematically fail to aim at E, G would not count as a φ -type agent.*

This constitutive end captures Bratman's suggestion discussed in Section 2.5.5. Recall his view that giving weight to certain ends can be definitive of a group and of a type of activity. For example, making legal chess moves with the aim of checkmating the opponent's king is such a definitive aim that is constitutive of playing chess. Similarly, having the end to walk, when hiking together, is constitutive of the activity of hiking. I want to emphasise that, here, I am not using the term 'constitutive' as it is used in the metaphysical literature on the constitutive ends of agency, such as self-intelligibility, autonomy or self-constitution. These ends, it is argued, are metaphysically required for agency as such. In this chapter, I will also not consider what kind of phenomenon gives rise to a constitutive end. This question will have to wait until Chapters 5 and 6, where I will argue that constitutive ends are part of certain social and legal practices. Practices that permeate the agency of group agents and turn them into agents of a certain type, with a characteristic practical identity.

Consider the following group agent with a constitutive maximizing aim E.

Definition 4.3.2. *Constitutive one-dimensional valuer*

Group agent X is a constitutive one-dimensional valuer with regard to its constitutive aim E iff

X has a normative perspective that is constituted by a web of shared policies of accepting weights regarding normative considerations such that these policies induce a ranking in which E has maximum weight relative to all other ends picked out by policies about weights that X accepts.

I will call such a group a constitutive ODV because its constitutive end always dominates and outweighs all other normative weights that guide its reasoning. This does not mean that the agent cannot consider the situation's other normative aspects, such as moral reasons to help, to keep promises or

to speak the truth. However, whenever these considerations conflict with the constitutive aim, the agent's accepted rankings of values always resolve this conflict in favour of the constitutive aim. This is a one-dimensional approach to conflict-resolution, hence the name.

With this said, I can now state my overall argument.

Possibility argument. Any valuing group agent with a *constitutive, maximizing, non-moral aim* is under rational pressure to be a *constitutive one-dimensional valuer*.

Responsibility argument. A constitutive one-dimensional valuer displays a functional and normative profile indicative of a lack of powers of reflective self-control, and, so, it is exempted from full participation in moral communities.

To establish the possibility argument, I have to show that A) there is rational pressure for agents to have shared policies about their maximizing, non-moral constitutive aim and that B) there is rational pressure to adopt a set of shared policies of acceptance about relative weights that give most weight to the constitutive aim.

Constitutive Aims and Group Deliberation

Take any group agent of type φ , G with a constitutive, maximizing, non-moral aim to E . The possibility argument follows from the shared policy and the definition of a constitutive aim. By definition, G 's having a constitutive aim to E means that G aims necessarily at E , and that if G would fail to aim at E , G would not be an agent of type φ . For the aim to E to appropriately guide the group's deliberation and reasoning – and thus to play its functional role – there must be a shared policy to give weight to E in all those situations in which E might play a role. Since G aims necessarily at E , the shared policy must cover all situations. Why is there this rational demand for such widely-applicable policies? One reason is that instrumental reasoning with regard to E depends on giving weight to E in deliberation. In shared instrumental reasoning, the members are committed to take the shared end as fixed and to infer the means for action implied by it.²⁶ Unless a shared policy to give weight to E is adopted by the group, G does not take E as fixed. In that case, G 's members cannot draw on the shared instrumental reasoning schema, even when they might believe on behalf of G that there is an open option that would be necessary for achieving E . Thus,

²⁶Bacharach et al. (2006) argues for this kind of instrumental collective reasoning.

any group G that does not adopt a policy to give weight to its constitutive aim E will by default be instrumentally irrational in those certain situations in which there are necessary means implied by E . The group will also fail to see these options as salient in deliberation. Thus, as a constitutive end, E would not appropriately guide G 's means-end reasoning, meaning that G will be systematically instrumentally irrational with regard to E , since it aims necessarily at E without intending the means.

But why should G care about this kind of instrumental irrationality? An answer to this question will have to appeal, in part, to the very nature of the constitutive aim and the reasons that ground it. I will explore this issue in the next two chapters. But there is also an agential aspect to this argument. Constitutive ends characterise groups. They can, as Bacharach et al. (2006) argues, prompt group identification. They can also influence membership deliberation by restricting membership to those who endorse the constitutive goal in the first place.²⁷ The relevant constitutive goals are also likely to have the same volitional necessity as Harry Frankfurt ascribes to certain ends, such that their absence would undermine the agent's self-governance. Self-governance is threatened partly because different ends imply different means and there is no way to resolve this conflict without giving up a constitutive end, hence the instrumental irrationality. This also means that there is no fact of the matter about where the agent stands on this issue, meaning that there is simply no self-governance with regard to this issue. Finally, a similar point, which concerns both the importance for group identification and self-governance, is made by Rovane (1997). Rovane highlights the importance of 'unifying projects' which can supply the agent with reasons for greater rational integration and coherence. Contrapositively, when a characteristic end does not play its functional role and one cannot count an agent as having it, its unifying force and normative pressure towards the point of view's greater coherence and integration is undermined.

These reasons explain the normative pressures towards avoiding systematic instrumental irrationality, which exerts normative pressure towards adopting the relevant shared policies giving weight to E . This concludes my argument for proposition A.

²⁷For example, suppose that the most important reasons for being a member of a football club are provided by the very activity that makes a football club what it is, playing football. Then, you will only be a member and considered as such if you are willing to actively participate in shared policies that give weight to the constitutive aims. Nobody wants to belong to a football club that doesn't play football!

The Normative Perspective of the ODV

How do we get from a shared policy to give weight to E to a shared policy about accepting relative weights such that E has the greatest weight in shared deliberation? Put differently, how do we show B) and, thereby, show that a group with a constitutive aim is under rational pressure to be a one-dimensional valuer? The first step is to remind ourselves of the role of shared policies of acceptance in Bratman's theory. Their purpose is to resolve conflicts among competing ends by ranking those ends according to the relative weights assigned to them.

The argument is a *reductio*. Let us assume that G is a valuing group agent with a normative perspective and a constitutive aim E which is non-moral and maximizing. Further, let us assume that G is not a one-dimensional valuer, but rather that its web of shared weight-giving policies over a range of situations does not mechanically give the most weight to E under every circumstance. We want to show that such an agent will undermine and jeopardise its own agency.

The argument is similar to the one given in the previous section. The fact that G is a valuing group agent means that, according to L&P's account, G is also an agent that can be the target of moral expectations, claims and obligations. As a morally capable agent with a normative perspective, it has adopted a set of shared policies that give weight to various moral considerations such as not doing harm, keeping promises and so on. These shared weight-giving policies allow the group to anticipate and respond to the set O , consisting of its obligations to the members of the moral community and their expectations of it. For G to respond to the moral reasons counting in favour of O – in other words, for G to endorse O and control its conduct in the light of the moral reasons supporting O – these moral reasons must be part of the shared moral policies. More importantly, as we can see from the previous section, for G to endorse O , these shared moral policies must, at least occasionally, be given maximum relative weight in deliberation, so that they are not always outweighed by the shared policies that *always* give some weight to the non-moral aim E in G 's shared deliberation. This means that shared policies of acceptance, which fix these weights, must, in relevant situations, give the most relative weight to moral reasons, at least in those cases in which these moral reasons support obligations that the group is held to by members of the moral community. It is the mark of a moral agent that moral reasons are sometimes such that certain actions are morally forbidden and some actions are obligatory.²⁸ Since we assumed that G is not a one-dimensional valuer, G will have a normative perspective in which it

²⁸Björnsson & Hess (2017) make the same point.

gives the most weight to moral considerations. Assume that it so happens that E systematically conflicts with the obligations in O . Maybe E is selling a harmful drug or tobacco, drilling for oil in fragile ecosystems or some other product the external – moral or more generally human – costs are not accounted for. It follows that, in these cases, regularly, when G faces a conflict between the necessary means implied by its self-interested, non-moral end E and the obligations O that it is held to, it will choose an element of O , since it gives most weight to it.

When instances of conflict abound, G will not count as a type φ agent, simply by definition 4.3.1. This means that G faces the kinds of threats to its agency outlined in the previous section. These will lead G 's members, as well as people interacting with G , to drop the intentional stance. This systematic irrationality is especially pronounced if E is a maximizing end. For example, consider the options open to an oil-producing firm. Suppose that it can choose between extracting oil from an environmentally unproblematic location in a desert or by means of an environmentally problematic, off-shore drilling operation off the coast of Brazil. In the latter case, the potential for profit is huge and the political situation is such that G will not suffer expected costs that would off-set these profits if something were to go wrong. A company which aims to maximise profits or its corporate value will believe that this aim will imply the necessary means to drill in both places. On the other hand, a company which has the simple aim of making profits by selling oil for money will be able to choose only one of the two options, and may be guided by moral obligations which forbid drilling off the coast of Brazil. This case illustrates why maximising aims put more pressure on the group to consistently and thoroughly adopt policies in its favour.

But we now arrive at the contradiction which we were after. We assumed that G is a valuing group agent with a constitutive non-moral maximising aim, but not a one-dimensional valuing agent. G 's normative multi-dimensionality implied that it must be capable of responding to moral reasons by adopting a suitable normative perspective. We can now see the deep tension between these two properties which has been created. If we assume that G is not a one-dimensional valuer, then, in order to satisfy the moral demands placed on it, G will have to act irrationally in a systematic way, which undermines its agency. If we assume that G is a well-functioning, internally-coherent agent that seeks to act rationally, we are forced to abandon the assumption that it can be a capable member in the realm of obligations, at least insofar as E systematically conflicts with the demands of morality. This means that either G does not meet the requirements of rationality or it responds to its rational pressures so as to become a one-dimensional valuer. A one-dimensional valuer is still a valuing group agent that meets L

&P's conditions of responsibility, but acting morally clearly requires it to act irrationally in a systematic way which, in turn, threatens its agency. But it seems inconsistent to hold an agent responsible for its immoral actions if these immoral actions are rational from its perspective. This, in a nutshell, is the challenge of the one-dimensional valuer.

Objections against the possibility of the ODV

There are three ways to respond to the challenge. First, one can question the existence of constitutive ends for groups or, second, question the functional role of the constitutive end that rationally induces the normative profile of an ODV. Alternatively, third, one can argue that a group with a constitutive end doesn't, in fact, meet the conditions for moral agency and, thus, that L&P's account of responsible agency is false.

I will address the first objection in Chapters 5 and 6. In the remainder of this chapter, I will consider the second objection and whether, by adopting a more Strawsonian understanding of holding responsible, we can avoid the conclusion that an agent for whom acting morally is irrational can still participate in moral relationships.

There are two ways to undermine the functional story which pushes a group agent to become a ODV. First, we might say that *G* reconsiders *E* when it realises that the necessary means implied by *E* conflict with its obligations and, instead, adopts the qualified end 'give most weight to *E*, unless *E* conflicts with the moral obligations *O'* to which we are held'.²⁹ If this is possible, *G*'s standpoint would be morally qualified and *G* would pass as a potential partner for interpersonal relationships in moral communities. I see two problems with this suggestion. The first is that it will not work with every constitutive end which is characteristic of groups of a certain type. In particular, it will not work if the characteristic end is immoral. Consider, for example, the end of a drug cartel which aims at making money by selling illegal drugs, participating in prostitution and extortion. It does not seem possible to qualify an end like this in a way which does not imply the cartel's dissolution or deeply irrational behaviour. Indeed, these kinds of groups are often outlawed, suggesting that the moral community does not believe that *G* can change to become a non-pathological group agent if *G* is a criminal organisation like, for example, a cartel or one of Germany's extremist right wing gangs. Here, the *can* is a rational can. One part of that rational pressure, as I have already mentioned in previous sections, also comes from reasoning related to membership conditions. The ease of creating or joining

²⁹This objection has been put to me by Herlinde Pauer-Studer and Niels de Haan.

new or different group agents with different constitutive ends, together with the fact that groups typically try to only include members who, in some way or another already endorse the group's constitutive end, hinders any process that turns the group into another type of agent – what's more, it increases the likelihood that members who disagree with the group's constitutive ends will simply exit the group. This also blocks potential objections to the rational processes which I have described as alluding to the idea that certain constitutive ends might block group self-identification, since, at least for some groups, these members have the option to leave the group and be replaced with members willing to adopt the group's point of view and allowing themselves to be guided by the group's constitutive aim.

But dissolving is not the only possible consequence of qualifying one's constitutive end in certain cases. Another possibility is that, at some point, the group no longer counts as an agent of the type characterised by the constitutive end and will, as a result, cease to exist as the agent which they were. Here, I am willing to concede that certain groups with non-moral ends may change their constitutive, non-moral end by reconsidering and qualifying it whenever it conflicts with the community's moral demands, thereby transcending their current group identity. But members will only be able to perform this task by acting outside their role and in conflict with the rational point of view of the group.³⁰ This could then amount to an a-rational change, a revolution or a putsch of the group's practical standpoint and normative perspective. Crucially, this change is a-rational (as well as irrational) from the perspective of the group: it is neither supported by its point of view or reasons, but by its members responding to their individual duties. Thus, this process is not governed by group-level reasoning, but by the strategic actions of individual members. Because of that, the resulting group agent is no longer identical with the previous one. This concludes my argument for my position that constitutive non-moral aims put a group which faces constant conflicts with the demands of morality under rational pressure to become a one-dimensional valuing agent.

4.4 The ODV and Responsibility

The ODV is a problem for L&P's account because it meets the conditions for being a valuing agent, and their conditions for being held responsible for a choice, while intuitively being a creature that cannot participate in

³⁰Maybe, as Stephanie Collins has suggested in conversation, people can even push the borders of their role and act morally within their role, thereby pushing the organisation in the direction of adopting morally upstanding shared policies.

arguments where certain justifications for actions – egoistic justifications, for example – do not work in moral discourse.

More specifically, I maintain that the ODV is incapable of a bad or a good will, where the agent can be held accountable responsible for its will. This is because it cannot act for the right reasons, because its subjective reasons are always geared to give the most weight to its constitutive end in critical cases where it conflicts with the demands of morality. Because it builds on a Davidsonian understanding of action, L&P's thin account of personhood doesn't capture this feature of ODV. Consider Gary Watson's remarks in an influential article on the connection between responsibility and an agent's will. The Hobbesian belief-desire account of action, Watson argues,

obscures the relevant distinction between [...] voluntary conduct and operant conditioning, between structural defects and virtues. [...] We can't be rightly blamed unless we have control of the causes of our conduct. If we lack control of our desires, as no Hobbesian view can preclude, we lack control over our wills.³¹

The ODV brings Watson's worry to light. L&P's account obscures the distinction between a group agent who values in the ordinary way and one with a structural defect such as the ODV. The same criticism applies to their thick account of personhood. As I argued in Section 2.5.3, that account may be understood to imply a hierarchical account of action. Hierarchical accounts of action face problems identifying the right kind of second-order endorsements which are 'special' in that they stand for the morally relevant aspect of an agent's will. The ODV is 'forced' to streamline its normative profile but does not voluntarily endorse doing so.

4.4.1 The ODV and the Powers of Self-Control

So, one problem that the ODV brings to the fore is that L&P's account of a valuing agent, which supposedly satisfies conditions of personhood, is compatible with the ODV's characteristic functional normative structures. The ODV is clearly not a competent participant in moral relationships understood in the Strawsonian way via the powers of reflective self-control. Although the ODV does form normative judgements by recognising morally salient features from its normative perspective, its normative perspective assigns the wrong weight to moral considerations, so that these considerations structurally and robustly fail to issue in moral permissions and obligations that restrict and guide the ODV's actions as a competent moral agent would

³¹(Watson 1996, 233)

be guided.³² Thus, the ODV is not just an immoral agent, but at least partially an amoral agent, because it lacks the deliberative capacity for a good will. Thus, the conformity of its will with the demands of morality depends on the contingent features of its circumstances, rather than on the role of its beliefs and judgements about the demands of morality. This means that the ODV does not meet the reason-comprehension condition and, as a result, does not meet the constitutive standards of a moral relationship.

Nevertheless, the ODV is answerable in a certain sense. It can, for example, identify the kind of ranking various normative considerations which have factored into its decision. However, it is not answerable in the sense that, as long as its deliberation is shaped by its constitutive end, it cannot be influenced by our moral arguments. Its justification should be read as explanatory and as stemming from its subjective, motivating ought. At best, we can use the agent's capacity for this kind of justification to count as evidence of it actually having a constitutive goal.

So, the first problem for L&P's account of responsibility is that the ODV would be wrongly assigned the status of personhood. This follows from the fact that L&P must argue that the ODV meets the conditions for responsibility. But, one might object that the ODV does not meet the normative judgment condition because it can't choose between, say, a means that is implied by its constitutive end and the obligations that it is under. This might be true, in that the ODV can't choose rationally between the two: rationally speaking, it ought to give most weight to its constitutive end, but, clearly, one can always choose something irrationally or pursue a 'perverse' desire. Moreover, the capacity to make normative judgements says nothing about what these judgments lead to, or even how the ODV is capable of choosing morally when doing so is consistent with its constitutive aim. So, under a normal reading of control that is relevant to responsibility, the ODV satisfies L&P's conditions of responsibility.

Thus, one way for L&P to avoid this conclusion is to understand their normative judgment condition as a condition that specifically concerns free choices over the group agent's own ends. This would exclude those agents whose unalterable ends or character traits exempt them from the practice of moral responsibility.

So, under a Strawsonian reading of responsibility and a modified reading of L&P's conditions of responsibility, the ODV is at least partially exempted from our practice of responsibility. Thus, it counts as an amoral agent. Nevertheless, the agent might be highly rational, capable of satisfying its consti-

³²Wolf (1987) also argues that the unalterability of certain ends or character traits does exempt agents from moral responsibility.

tutive end and capable of diachronically extended activities. These capacities might still leave room for certain kinds of interactions, albeit not those interactions characterised by the importance of moral argument and the exchange of moral reasons. Instead, these interactions are shaped by aretaic appraisals which have the function of regulating, controlling and managing the ODV's behaviour by supplying it with self-interested reasons to adhere to the norms important to our society. Thus, our treatment of the ODV group agent might resemble our treatment of a psychopath who does possess the requisite rational capacities not to act on one-off impulses, but who is incapable of caring for the interests of others for some other structural reasons.

The ODV is distinguished from an agent that can be held responsible but acts immorally by the fact that an immoral agent has no constitutive non-moral aim. It has a moral end, or at least a rational way towards forming such ends, such as an NGO, but then adopts bad ends or purposefully decides to violate its obligations. For example, an NGO might embezzle money or overpay its employees relative to its material commitments to its stated goals, where such actions are in no way rationally implied by its constitutive aim. A group agent might also be considered immoral rather than amoral if its constitutive aim never conflicts with the demands of morality, but the group nevertheless does not comply with the moral demands placed upon it. For example, consider a sports club. If the sports club discriminates against people or charges high fees of its poor fans, even though its finances are secure, then it can be blamed for both of these actions. It displays a bad or ill will, which makes it an immoral agent rather than an amoral agent.

4.5 How do the two Challenges Hang Together?

The two challenges I have presented both bear on the will of a group agent. The misappropriation case highlights how difficult it is for group agency accounts to distinguish between group actions and group activities in those cases in which the group is not in control due to a desire or a goal that is 'external' to the group's mind. The case of the ODV, on the other hand, shows that group agency accounts have a blind spot covering the character or structural aspects of a group's normative profile that can impact their capacity to act for a bad as well as a good will.

Both challenges concern what one might call *unusual* ends: both involve a kind of necessity. In the misappropriation case, one solution to the problem of external ends – in the sense of ends undermining agential authority – involves the claim that certain ends have *volitional necessity* for the agent, such as the end of pursuing one's own welfare which didn't play its functional

role in the misappropriation case. Because Sue and Bob's selfish decision on behalf of the group were in conflict with it, their choice was not an instance of the group's self-governance. Rather, it was merely an instance of a group activity.

On the other hand, in the case of the ODV, the unalterability of a non-moral maximizing end has the opposite effect on the agent by depriving it of its capacity to be a part of the moral community. Importantly, the problem is not that the ODV's actions are not expressive of the kind of agent that it is, which was the problem in the misappropriation case. Rather, because its end is insensitive to the morally salient features of its environment, the ODV is stripped of some of the powers of reflective self-control. One might think that the external or unchangeable nature of the constitutive end also diminishes self-governance by the ODV. In the misappropriation case, we tried to show that, when the constitutive end is not endorsed, the group agent's agency is under threat. The ODV's constitutive end, as I have argued above, is relevant for self-governance, but is an end of the wrong sort, morally speaking. It only shows that certain constitutive ends do not get you the kind of morally salient self-governance and control, while we do need some kind of constitutive end to constrain that set of member-made decisions which can count as the group's decision. This tension brings another aspect to the surface, showing how they are essentially two-faced. We want them to fulfill a certain function, but require them to be capable of interacting with us, ideally in a moral way.

This brings me to an interesting conclusion for group agency accounts in general. On one hand, these accounts will require some role for a practical identity that is fixed so as to provide us with a principled way of distinguishing members who interpret their role in a rogue way from those whose contribution ensures the group's agential authority. But there is a danger looming. Not just any practical group identity will provide us with an agent capable of moral agency. Certain non-moral maximizing aims will cripple and undermine the agent's cognitive capacities in a way that exempts them from our reactive attitudes and our moral relationships. This does not mean that we can't punish these agents or end their existence. It just means that, when applied to these agents, our reactive attitudes and practices of responsibility will lose their normal expressive function.

4.6 Summing up

Often times, when we talk about groups, we represent them as means to certain ends. For some groups this relation might be captured by modelling

the group as an agential system with a constitutive, non-moral maximizing aim. In this chapter I have argued that these groups are under rational pressure to become one-dimensional valuers. I have also adopted a common understanding of responsibility that is situated within a largely Strawsonian framework of moral responsibility. In it I distinguish between two ways of holding responsible: Accountability-responsibility tracks the Wallace's powers of reflective self-control that are necessary to participate in moral relationships. Aretaic-responsibility captures appraisals and responsibility ascriptions that concern the character of an agent. I have argued that the one-dimensional valuing agent is not the appropriate target of accountability-responsibility, but can participate in relationships which are characterised by aretaic-responsibility appraisals.

5 A Constitutive Account of Group Agency

5.1 Introduction

In the previous two chapters, I developed two challenges for theories of group agency. Both drew on situations in which a group agent doesn't find itself in an action entirely of its own making. Rather, in those situations, it is a co-author in a script with a framework fixed by the social practices and institutions which characterise the place it inhabits.¹ It is a pervasive characteristic of sophisticated societies like our own that we bring groups into existence to solve problems. In principle, this is unproblematic. Corporations, for example, are entities who aim to produce and sell goods or services efficiently and to do so while meeting existing demand. These two functions are nicely captured by the concept of profit. For a firm to make profit, it must produce something that is in demand, which it can sell for a price that at least covers its costs. A competitive environment ensures that this happens efficiently, as happens when a company makes a profit. Thus, under ideal circumstances, a corporation makes profits if it efficiently meets a demand, where the costs of the product are included in the price at which it is sold. Similar stories can be told for NGOs, sports clubs or government agencies. Importantly, in our society, all those groups are assigned variously precise functions with correlative goals and success conditions. If we keep this in mind, it does not seem unusual to treat groups as means,² to allow them to be owned, bought or sold or to be taken out of existence when they are no longer needed or fail to fulfil their role. Note, for contrast, that even though a person might be assigned a certain function, status or role, the person's *value* is in no sense restricted to the successful fulfilment of this function. There is nothing, in principle, in the person's physical and mental constitution that prevents her from shedding a role or indulging in a new pursuit which is at

¹MacIntyre (2013, 213)

²See Albertzart (2017), who argues that corporations are mere instruments without the capacity for moral agency and Wringer (2014) who investigates the implications of treating groups as means.

odds with the contingent external factors and or with her own abilities. For groups, artificial constructs with ‘bodies and souls’ which are designed and shaped intentionally by designers with a purpose in mind, the same is not at all obvious.

Nevertheless, based on the previous chapters, it seems also likely that groups may introduce a culture of deliberation and of giving weight to normative considerations that can emulate and fulfil what Wallace has called the powers of reflective self-control. The task for a theory of group agency, therefore, must be to do justice to this dual nature of groups and to provide a framework in which we can explicate and make transparent both faces of groups: on one hand, their capacity to act morally and perhaps autonomously and, on the other, their potential function as tools. Ignoring one of them might be intellectually tidy, but can also lead us astray normatively.³

In this chapter, I set out to provide just such an account. It is a collective variant of the constitutive theory for individual agency that David Velleman has developed over a series of articles and books. In many ways, it offers a natural extension of Philip Pettit’s understanding of group self-identification, but does so without the metaphysical assumption that group attitudes supervene on the attitudes of its members.

I will proceed as follows. In Section 5.2, I describe and motivate Velleman’s constitutive theory of agency. In Section 5.3, I develop an analogous account for a theory of group agency in which the constitutive aim of autonomous group actions is group self-intelligibility. I lay out central aspects of that theory by specifying the concept of making sense from the perspective of the group (Section 5.4), the essentially shared nature of group self-intelligibility (Section 5.4.1), the nature and grounds of a group’s practical identity and their influence on a group agent’s reasons for action (Section 5.5). I conclude this chapter by investigating how well the concept of group self-intelligibility captures the idea of an autonomous group will and how the constitutive account can help us deal with the two challenges from chapter 3 and 4 by exploring the consequences for our practice of responsibility (Section 5.7).

5.2 A Constitutivist Theory of Individual Agency

I want to extend L&P’s existing theory of group agency to capture the idea of full-fledged autonomous intentional agency, while also leaving room for interpreting groups as instruments finely tuned to pursue a certain aim. To

³This point has been ardently made by Kusch (2014).

do this, I will, for now, set aside the instrumental face of group agents and focus on groups as autonomous agents. I propose that we approach this question by looking at what the very notion of autonomous agency entails.

This project need not start from scratch. In locating autonomy, ethics and the normativity of practical reason in the nature of action and agency, I follow a venerable tradition in the philosophy of action and agency originating in Kantian philosophy of action. Philosophers who take this approach ‘seek to show that being subject to the authority of certain norms or principles is somehow constitutive of agency.’⁴ They try to ground the normative force of moral requirements not in contingent things like desires and inclinations, but in the universal concept of action and agency.⁵ These philosophers are called *constitutivists*. They do not focus on the less developed, ‘lower’ forms of agency exhibited by animals, robots or children, but rather on full-fledged, autonomous human agency that, when exercised, expresses a good will.

So, maybe we can locate something akin to full-fledged group agency by studying what constitutivists take to be constitutive principles of autonomous agency. If we find that some valuing group agents have the capacities to satisfy the constitutive principles, then we might finally have an account of group agency which robustly satisfies the conditions of moral responsibility such as the power for reflective self-control that we have analysed earlier, while also excluding activities of the sort found in the misappropriation case and agents such as the one-dimensional valuer as not fully autonomous agents.⁶

In what follows, I will be interested in a specific kind of constitutivism that has been defended by Velleman (2009*a,b*, 2006). Velleman’s account is of particular interest for several reasons. First, it is, as Velleman claims, highly compatible with Daniel Dennett’s account of the intentional stance and Dennett’s conception of the self as a ‘center of narrative gravity’, thereby allowing for a natural extension within our general metaphysical framework.⁷ Second, his theory of agency is not as normatively committal as one of the main alternative constitutivist theories developed by Korsgaard (1996, 2003).⁸ Finally, one of Velleman’s central motivations is to provide a criterion for systematically distinguishing intentional action proper from between mere human activity that is subconscious or not properly endorsed. Since my aim is to

⁴Silverstein (2015, 1127).

⁵See Tubert (2010).

⁶To recap, the power for reflective self-control consists of both the power to grasp, understand and apply moral reasons and the power to act in light of such reasons.

⁷See Velleman (2006, Chapter 9).

⁸See citarruda2016constitutivism for an argument showing that Velleman’s theory is less normatively demanding than Korsgaard’s.

provide such a criterion for group cases such as the misappropriation case, Velleman's account offers an obvious solution.

Velleman develops his constitutive theory of agency with two goals in mind. First, he wants to distinguish between intentional autonomous action and activity.⁹ Second, in developing his theory of agency and intentional action, he wants to defend an internalist conception of practical reason which nevertheless satisfies certain universal standards.

The cases in which Velleman is interested, and which he wants to distinguish from full blown, autonomous intentional action, can be described as purposive activities. Purposive activities can be explained or rationalised by identifying a Davidsonian belief and desire pair that causes them. What distinguishes actions of this kind from autonomous intentional actions is that they lack the relevant 'causally normal' way in which this pair influences, guides and controls the agent's behaviour. Here is one such case.

Suppose that I have a long-anticipated meeting with an old friend for the purpose of resolving some minor difference; but that as we talk, his offhand comments provoke me to raise my voice in progressively sharper replies, until we part in anger. Later reflection leads me to realize that accumulated grievances had crystallized in my mind, during the weeks before our meeting, into a resolution to sever our friendship over the matter at hand, and that this resolution is what gave the hurtful edge to my remarks. In short, I may conclude that desires of mine caused a decision, which in turn caused the corresponding behavior; and I may acknowledge that these mental states were thereby exerting their normal motivational force, unabated by any strange perturbation or compulsion. But do I necessarily think that I made the decision or that I executed it? [...] Indeed, viewing the decision as directly motivated by my desires, and my behavior as directly governed by the decision, is precisely what leads to the thought that as my words became more shrill, it was my resentment speaking, not I.¹⁰

⁹Although his aim is to describe the phenomenon of the exercise of human agency, Velleman (2009a, footnote 1) does extend that phenomenon to include 'action on the part of nonhuman organisms', namely action as 'behavior controlled by the organism, not just by one of its constituent subsystems or parts.'

¹⁰Velleman (2009b, 127). Also note that (Velleman 2009b, footnote 193) does not say that the agent isn't absolved of responsibility in every activity of this kind. His general ability to interpose or prevent the activity in question establishes his responsibility. I am not sure whether Velleman's analysis is correct, since, if the activity were up to me, then it seems to have been somehow under my control after all. I do agree, however, that an

What is abnormal about these cases, Velleman claims, is not the causally efficacious role of the relevant mental states in the activity. Rather, it is that the agent herself is not part of how they come about and how they are realised behaviourally.

One solution to this problem has been proposed by Frankfurt (2018), who argues that autonomous intentional actions are not just caused by Davidsonian belief-desire pairs, but by those desires of which the agent is aware and that she desires to have. This is the hierarchical model: it is not enough for the agent to be compelled by her motives, but she must also have a second-order desire to consciously favour being compelled by the first-order desires which she enacts. This rules out characterising as intentional the sub-conscious purposive actions, such as Velleman's story about the sub-conscious break-up with a friend, since, in this story, the agent wasn't aware of his desire and decision to break up with his friend.

Frankfurt's solution runs into a different problem. Velleman (2009b, 13) argues that

[t]he hierarchical model [...] doesn't distinguish sufficiently among the subject's responses to his own motivation. A favorable disposition toward his first-order motives automatically implicates the subject in causing their behavioral results, according to this model, irrespective of why or how he is favorably disposed toward those motives. It doesn't matter, in the hierarchical model, whether the subject is satisfied with his first-order motives because of depression or boredom or laziness – or, alternatively, because he is responding to their force as reasons.

Thus, the problem with Frankfurt's proposal is that he treats as intentional actions which are done for first-order motives but supported by second-order motives from which the agent might be alienated.¹¹

Velleman's solution is that in intentional, autonomous action proper, the agent's behaviour is not just governed by his relevant beliefs and desires, which are supported by certain second-order desires. Rather, he argues that there is a single special second-order desire which allows the agent to treat considerations as reasons, beyond just being favourably disposed towards them. This special second-order desire is the desire for *self-intelligibility* or

absolute inability to intervene and to control my action would render me unfit as a proper target of the reactive attitudes.

¹¹This is also the reason why Philip Pettit's account of group identification does not succeed at ruling out the possibility of the one-dimensional valuer.

self-understanding.¹² This aim may be sub-agential, meaning that it regulates our practical reasoning without being explicitly represented in it.¹³ Whereas other second-order aims might express the agent's endorsement of first-order desires without these endorsements being perceived as reasons by the agent, self-intelligibility is, so to speak, the ultimate second-order aim where endorsement and the 'taking as reason' coincide. An agent's aim of 'making sense of himself to himself' not only governs and influences how his desires and beliefs are realised. It also 'determine[s] which [desire] prevails, by adding its force to the one that strikes him as making most sense for his character to manifest.'¹⁴ In doing so, the aim plays a 'modulating role in our practical affairs. It influences which desired objects we choose to pursue, how we harmonize them with one another, organize our efforts toward them, and express our thoughts and feelings along the way.'¹⁵ What makes sense is a product of the agent's motives, attitudes, character traits and the circumstances in which the agent finds himself within the context of folk psychology. This way of understanding sense-making is supposed to be non-normative: it does not amount to the question of what the agent ought to do, but is concerned only with what makes sense given the backdrop of one's attitudes, character traits and dispositions.¹⁶ Understood in light of this standard of success for intentional actions (where the standard of success is the end of self-intelligibility), *practical reasons* are considerations in light of which a certain action makes sense. When an agent *chooses* rationally, he imagines or entertains different alternatives open to him and settles on that alternative which makes the most sense to him. In settling on a certain narrative, he comes to believe that this narrative will come true. Here, again, his aim of self-understanding comes in: it compels him to enact the narrative which he has settled on, in order for the narrative to come true and thereby afford him the relevant self-understanding. He gains self-understanding in the sense that, as a rational agent, he has settled on the narrative which makes most sense to him. 'We are thus in a position to make up our forthcoming behavior,' Velleman claims, '[m]aking up what we will do is, in fact, our way

¹²Velleman's account of the constitutive aim of action has changed over time. Whereas Velleman (2009b, Chapter 7) identified the constitutive aim of action straightforwardly with autonomy, on his most current account, autonomy, as a property of actions, is only a derivative feature of behaviour which aims at self-intelligibility. See Velleman (2009a, 130, 132-33).

¹³See Velleman (2009b, footnote 29).

¹⁴See Velleman (2009a, 15).

¹⁵See Velleman (2009a, 28).

¹⁶See Velleman (2009a, 13, 16). Similarly, the self-conception by which the agent makes sense of himself to himself 'is not evaluative: it is a description under which your actions and reactions make sense to you in causal explanatory terms.'

of making up our minds to do it.’¹⁷

To be capable of aiming at self-understanding, the agent must be, on one hand, *self-aware* of her relevant intentional attitudes so that they may play their characteristic role in narratives that make sense of herself to herself. On the other hand, she must also be capable of *theoretical inquiry* in order to be able to judge whether the set of considerations which she weaves together to form a narrative makes sense compared to a different narrative of the same action or compared to a different action endorsed by a different narrative.¹⁸

To better grasp the concept of self-intelligibility, let us return to Velleman’s objection against Frankfurt. We can now see that self-intelligibility is both absent from the decision to choose that narrative which makes most sense to the agent – for example, to solve the conflict with his friend in a constructive manner – and subsequently from the way that this decision is enacted. In lacking the aim of self-intelligibility, these cases lack a story about the agent’s influence on controlling her own behaviour. Finally, self-intelligibility is constitutive of autonomous intentional action in that you cannot fail to care about it. It sets an inescapable normative standard for you. You might be able to suspend it temporarily, perhaps by going to sleep or become intoxicated. But, due to your drive towards theoretical inquiry and your awareness of the world and yourself in it, that standard will bring you to make sense of the world and yourself and, so, to understand yourself. It is also inescapable in another sense: were you to ask whether and why you should care about understanding yourself, you would instinctively try to solve this problem by reasoning practically about it. The twist is that such reasoning presupposes the very end which you are trying to argue for, thus making the question moot.¹⁹

Whatever the success of Velleman’s story about intentional action, it has also met some criticism. It has been argued, even by Velleman himself, that his theory is intellectualist and runs into familiar problems faced by cognitivist theories about intention.²⁰ The overarching idea of cognitivist

¹⁷Velleman (2009b, 24). Understanding intentional actions via Velleman’s aim of self-intelligibility affords a compelling explanation of a famous insight by Anscombe (2000). She has argued that if we intentionally *A*, we know that we are *A*ing, and this knowledge is immediate. This is clearly true if we understand intentions as those beliefs which make most sense to us such that, by enacting them, we satisfy the goal of being intelligible to ourselves. In groups, immediate self-knowledge of one’s actions is much less intuitive, because the group’s point of view might not be as transparent to us as an individual agent’s point of view is to that individual agent.

¹⁸(Velleman 2009a, 17, 19–20).

¹⁹Velleman (2009a, 137)

²⁰See Velleman (2009b, 31) or Bratman (2018, 45) who calls the view, tongue in cheek, Velleman’s intellectualism.

theories of intention is to ground the norms of practical rationality governing intentions in norms of theoretical rationality, which is typically done by postulating some relation between intention and belief. In Velleman's case, intention is a special kind of belief, but weaker assumptions are also possible. Although I don't believe that my arguments depend on the subtle differences between cognitivism and non-cognitivism, I will return to this issue in Section 5.6.3. In this chapter, I will simply build on Velleman's fruitful idea that self-intelligibility is constitutive of autonomous action and explore its potential application to cases of group agency.

5.3 A Constitutive Theory of Group Agency

Intuitively, a group aims at self-intelligibility when its behaviour is guided by a shared narrative or at least a narrative that the group collectively accepts in that it makes most sense to *itself*. But how can a narrative possibly make sense *to* a group's self? What is this 'group self' to which certain considerations make sense while others don't? According to Velleman, the "'self'" expresses the reflexivity of representations, especially their notional reflexivity, the property they possess when they represent their object as their subject. [...] We use "self" to report a thought or utterance containing "I" just as we use "present" to report a thought or utterance containing "now".²¹ This idea is also contained in L&P's claim that using the first-person plural pronoun allows members to speak with authority about what the group thinks and does. They argue that using the first-person plural pronoun requires identifying with the group, which, as we have seen, in turn requires adopting the group's point of view. So, we can reformulate the idea of making sense to the group's self as making sense of an action *from the group's point of view*.²² Group reasons, then, are considerations in light of which the group's doing that thing would make sense from the group's perspective.²³

²¹See Velleman (2006, 354).

²²In this section, I will be using the expressions 'point of view' and 'perspective' interchangeably. Alternatively, we could understand 'making sense of an action to ourselves' as a narrative that we could accept or 'collectively understand' and, by aiming at self-intelligibility, subsequently enact. I am not sure whether using the idea of a collectively accepted narrative would yield a different theory. I think that this idea lacks the perspectival feature which we get when we talk about an agent's point of view or standpoint. This perspectival feature better captures the idea of a singular locus of agency which is not contained in the idea of a shared narrative.

²³This emphasis on the perspective for understanding self-intelligibility is in line with Bernhard Schmid's account of plural self-awareness, which we have considered in Section 2.4.1. That account also involves adopting the group's perspective, normative pressure

5.3.1 Group Self-intelligibility and the Point of View

How should we understand the group's point of view, from which a member (or a set of members) makes sense of an action? By what principle can we distinguish between elements that are part of that point of view and those which are not and thereby do not function as group reasons? Above, I wrote that the group point of view is that point from which the group member (or set of members – I will oftentimes leave out this plural possibility) asks 'What would make most sense for *us* to do?' In asking this question, she projects herself into the rational or narrative space of the group agent, a projection which entails framing the decision situation as one where the locus of agential control lies with the group. This means that the individual thinks of the decision situation as defined by the group's choices and causal actions, rather than her own choices causally independent from the choices of others. This parallels Bacharach et al. (2006)'s understanding of agential transformation from the 'I' mode to the 'we' mode in his theory of team reasoning and L&P's understanding of group identification. Understood this way group reasons must be considerations which apply to causal choices that are open to the group and depend on its abilities.

We can continue to follow L&P in thinking of the point of view of the group as consisting of the commitments and ends which the group has previously decided on. Thus, a group member's projecting herself into the narrative space of the group in order to figure out what would be intelligible to it means that she takes into account the mental states and commitments generated by the organisational structure, comprising the hierarchy and the relevant decision mechanisms. The elements in the narrative space are held together by rational requirements such as consistency, completeness, means-end coherence and other requirements which govern how beliefs, ends and intentions fit together, as well as what kind of patterns of mental states are coherent for an agent to have over time. These requirements stem from our folk psychological understanding of the different intentional states and are part of applying the intentional stance. They filter out certain considerations that would otherwise count as reasons for group actions. For example, they filter out considerations that support an intention inconsistent with the group's other commitments. They also highlight salient considerations like the necessary means for achieving the ends of existing intentions. Further, they provide the group with reasons to stick with an intention instead of

towards unification of the collective point of view and ownership of the actions by the collective. Since Velleman (2009b, 325) argues that '[t]he will emerges as a by-product of curiosity plus self-awareness', we follow a common strategy in highlighting the perspectival nature of self-intelligibility.

abandoning it without cause. This means that we should think of the perspective of the group not only as a narrative space, but also as a rational space. In section 5.4.1 I will consider different ways by which a group can ensure satisfaction of these rational requirements.

But, at this point, my account of group agency departs from proceduralist accounts. These accounts faced problems in that they could not, in a principled way, distinguish between considerations that were genuine group reasons and considerations in which the group played no role. A case of the latter kind was one in which people abused their roles within a hierarchy to misappropriate funds belonging to the group agent, for their own personal benefit and the group's detriment. How do we establish, in a principled way, whether a consideration such as an end belongs to the group's point of view and can function as a group reason which gives rise to and justifies an autonomous group action? In the individual case, the point of view does not only exclude certain elements because they would lead to irrational patterns together with the other mental states, but also those elements that are incompatible with the agent's practical identity or self-conception. This suggests that we can appeal to the practical identity of the group to exclude considerations in favour of self-dealing as failing to rise to the level of group reasons.

This approach is in the spirit of Pauer-Studer's constitutive account of group agency. According to her, '[t]he constitution and the self-definition of a group agent provide the normative standards by which to assess the behavior of the group agent in terms of coherence and intelligibility.'²⁴ But

²⁴See Pauer-Studer (2014, 1635-6). The idea of using a practical identity to delineate between aims that are external to an agent and aims that are internal has also been appealed to by Frankfurt & Frankfurt (1999, 132, 135), who writes that a 'person acts autonomously only when his volitions derive from the essential character of his will. [...] Even though a person's interests are contingent, they can belong to the essential nature of his will.' A few years earlier, Rovane's influential work on personhood makes very much the same claim, although (Rovane 1997, 173) speaks of unifying projects.

And the function of such a unifying project is to provide a perspective from which the group can deliberate and act together – that is, as one person with one rational point of view. [...] they provide the principle by which intentional episodes are distinguished as belonging to a particular group person's rational point of view, for it is in the nature of such a project to dictate how the efforts of a group of human beings must be coordinated so as to carry it out.

Miller's 2001, 2003, 2006 teleological account of social institutions also appeals to the idea of institutional ends to discern actions on behalf of an institution from mere group activities. Due to diverging ontological commitments (he endorses a strong kind of individualism) I will not pursue his account any further.

what exactly is the practical identity of an agent? Following Velleman, the practical identity or the self-image of a person ‘embodies a person’s sense of self. [...] it is the representation of a person considered non-first-personally but identified as the subject by some other, extrinsic means.’²⁵

Now, consider Velleman’s comparison of his view to how Christine Korsgaard understands an agent’s practical identity in her influential work *The Sources of Normativity*.

Because an agent identifies with the principle that dictates his choice, the principle can actually be expressed as a self-conception:

An agent might think of herself as a Citizen of the Kingdom of Ends. Or she might think of herself as someone’s friend or lover, or as a member of a family or an ethnic group or a nation. She might think of herself as the steward of her own interests, and then she will be an egoist. [...] And how she thinks of herself will determine whether it is the law of the Kingdom of Ends, or the law of some smaller group, or the law of egoism, or the law of the wanton that will be the law that she is to herself.²⁶

The law with which the agent identifies, by adopting one of these self-conceptions, is in fact a principle of choosing to act in particular ways under particular circumstances – circumstances that the law constitutes as reasons for acting in those ways.²⁷

Identifying with a principle still allows the agent to abandon these principles whenever she sees fit. Thus, these principles of choice are still too easily alterable. For that reason, Korsgaard qualifies her characterisation of these principles as follows.

Our contingent practical identities are, to some extent, given to us – by our cultures, by our societies and their role structures, by the accidents of birth, and by our natural abilities. [...] Although I have been suggesting that we do make an active contribution to our practical identities and the impulses that arise from them, it remains true that at the moment of action these impulses are the incentives, the passively confronted material upon which the active will operates.²⁸

²⁵Velleman (2006, 6)

²⁶Korsgaard (1996, 101)

²⁷Velleman (2006, 297)

²⁸Korsgaard (1996, 239-240)

I will be adopting a conception of practical identities that lies in the intersection of Velleman's characterisation as being external to a person and Korsgaard's view that they take the form of principles of choice.²⁹ This hybrid approach means that these principles of choice are not up to the group agent, although it is, in most cases, up to the agent how they are realised and how they govern its actions. In Section 5.5.1, I will dive deeper into the reasons why I believe that some parts of a group's practical identity must not be up to the group agent and what I consider to be the social and ontological status of those parts.

5.4 Group self-intelligibility

Let us now take stock. Let G be a group and M be a non-empty set of members of G and X be a choice open to G .

Definition 5.4.1. Basic Group Self-intelligibility: *M makes sense of X from G 's perspective iff the relevant members take G to be the causally efficacious unit with regard to X (and any non- X alternatives) and M would be able to take X to be explained by G 's practical identity, its circumstances and its other mental states, where these elements constitute G 's sufficiently shared point of view.*

If M aims at making sense of the group's behavior to the group, M will then arrive at a narrative of the group's action which makes use of folk psychological terms, thereby 'making up the mind' of the group. This will bring each individual member to do her part in enacting the group's decision. Of course, just as in the individual case, the aim of making sense does not need to be represented by the members individually or collectively, which is why the condition of sense-making is a *counterfactual* one. Often, it is intelligible to simply enact a previous decision or a policy that has been decided on. The aim of self-intelligibility only arises when the group must choose among competing ends or about how to implement a specific aim. Only when individual members somehow become aware that the normal functioning of agency isn't ensured do they, as a sub-group or as an individual member, ask themselves what action would make sense *to us*. In this way, the aim of self-intelligibility structures the group's point of view unlike simple

²⁹Perhaps that isn't true for certain abilities which surely also constitute a group's practical ability, such as having some kind of equipment, building cars or drilling for oil. But it definitely rings true for ends which are contained in the principle of shareholder governance and other legal norms governing the practical identity of a corporation.

group (self-)identification with the group. Self-intelligibility aims at a greater overall coherence by grounding the requirements of rationality.

Group self-intelligibility also plays the same role as in Velleman's individualist account of self-intelligibility, in that it is the standard by which we judge whether a group activity reaches the level of autonomous intentional action by telling us which considerations count as group reasons and which do not.

There are three parts in the definition that require further elaboration. The first and second notions concern who the *relevant* members are and how they *relate* to the point of view. What does it mean that the elements of the shared narrative of the group's action must be sufficiently *shared*, not only among a sub-group, but also among individual members deciding or acting on behalf of the group agent (or the sub-group) and the rest of the group agent? The notion demanding the most care concerns the group agent's *practical identity*, the self-conception of the group that the members represent reflexively when they think of who they are as one group agent among other groups. Here, I propose a novel way of understanding the ontology of a group's practical identity. In that context, I will provide another argument to the effect that this identity must be at least partially fixed from the outset.

5.4.1 The essential sharedness of group self-intelligibility

Developing a narrative that makes sense from the group's point of view is not something that any member can do on their own. Rather, it depends in two ways on the social web that makes up the group's organisational structure. In the following pages, I explore how the process of making sense is essentially a social endeavour. It gains this social dimension because of the process of determining the relevant members that have to make sense from the point of view, as well as how they relate to it.

The Organisational Condition on Group Self-Intelligibility

In identifying the relevant members who must aim at group self-intelligibility for the group to count as self-intelligible, I will build on L&P account which I have explored in Chapter 2. I will be focusing on the organisational hierarchy, which I will understand via formal and informal policies of procedure – shared policies which pick out who is to decide on a certain issue on behalf of the group and how they are to do this, for example, by means of deliberation, majority vote or transferring a decision to an expert – to determine a group's attitude on a certain issue.³⁰ In doing this, I also retain L&P's

³⁰See Section 2.4, where I explore the concept of an aggregation function at length.

idea that whether the intentional states which are thereby picked out satisfy certain rationality assumptions depends on the kind of procedure at play. Consequently, this means that a simple majority vote on logically connected issues does not allow the group to achieve robust group self-intelligibility.

Adopting this framework allows us to attribute self-intelligibility to a group, even if some or even most of its members are alienated from it. This is true whenever the alienated members are either not relevant to a substantive part of the point of view of the group or to the workings of a specific policy of procedure, as, for example, they would be when their input to the specific policy of procedure does not determine the output. Imagine a situation in which a sub-group must deliberate about a decision, while one member is alienated from the group. This sub-group might still achieve group self-intelligibility as long as that alienated member doesn't undermine the result of the shared deliberation. A group action might also count as group self-intelligible despite the alienation of certain enactors as long as, due to the influence of relevant incentive mechanisms, they contribute to realising the group's decision in a way that is in line with the group's point of view.

Similarly, a group counts as aiming at self-intelligibility with regard to a certain choice X even if large parts of that group are unaware that some authorised subset of the group has decided on behalf of G whether or not to X . This is the norm in big corporations in which decisions by, say, the marketing department are not necessarily known by those who participate in production. This permits a more nuanced picture of self-knowledge. Even though, in acting, a group agent gains self-knowledge through collective self-awareness, this collective self-awareness will be limited to those members who decide on behalf of the group and those who enact the resulting decision. This kind of group self-knowledge is not immediately accessible to other parts of the group that might need it. This is mainly a question of degree, since other parts of the group can gain that self-knowledge by means of inquiry. This is not unlike the individual case, where, even though you might have had direct access to an action which you did, as that action recedes into the past, your memories fade to the point of being unaware of why you did what you did.

Group Self-Inquiry

The organisational condition highlights that a group agent can act in a self-intelligible and thereby autonomous way, even if some or even most of its members are alienated from the decision and the resulting action. Furthermore, certain group decisions which are made in a group self-intelligible way will not be known to other members. Indeed, for most large groups like corporations, states and non-governmental organisations, there will be no single

member who is aware of all of the decisions made on behalf of the group in a self-intelligible way. Therefore, any point of view from which members decide on behalf of the group will necessarily be gappy. What follows from an excessively gappy point of view will be underdetermined, meaning that individual members might arrive at different conclusions or even at conclusions incompatible with the complete point of view. So, choosing and acting on behalf of the group in a group self-intelligible way must navigate between two horns: the practical impossibility to know every aspect of the point of view of the group (a feat also out of reach in the individual case, as we regularly forget commitments made and epistemic conclusions drawn) and a gappy point of view that hamstrings any robust way of making sense from the perspective of the group.³¹ Thus, we must say something about how members deliberate on behalf of the group or about how they *group self-inquire* – how they close the relevant gaps in their perceived point of view of the group when coming up with a solution to a practical problem.

To solve this problem, I will again avail myself of the tools I have used earlier to analyse some of the intricacies of L&P's aggregation functions: Bratman's shared policies of giving weight to normative considerations. These policies together with the aforementioned policies regarding procedures shape the role of each member within the group. They do so by forming the normative framework within which each member reasons about the practical issues that the group confronts.

Shared policies of giving weights specify, for each member, which previous decisions regarding intentions, goals and preferences, and which beliefs and values to consider when deliberating and reasoning about what to do on the group's behalf. Thus, a point of view does not only consist of the group's mental states – such as group goals, preferences or beliefs – and the practical identity of the group, but also of shared policies about giving weight that pick out, for each member in their respective role, which of these elements of the group's mind are salient. This may also mean that these shared policies give different weight to the same considerations for group members playing different roles. For example, environmental concerns might play a greater role for those corporate employees who must pick a supplier for a certain component used in the production process, while these concerns play only a marginal role for the human resource department.

It is important to see that the point of view so conceived does not need to determine the outcome of the kind of shared deliberation that it makes

³¹I think that this is one of the most complicated issues in applying any account of group action and one that seems to be often overlooked. For example, L&P seem to assume that each member has complete knowledge of the point of view of the group agent of which they are a member.

possible. For example, a corporation might have formed the commitment to value environmental concerns, in addition to its commitment to making profits. If it has not specified how these considerations should be weighed, then mid-level managers in the hierarchy will occasionally be faced with difficult decisions, such as whether to conduct drilling in a specific cost-effective, but risky way. The mid-level manager's decision might then turn out to violate the commitment that the board has made regarding environmental concerns. But, in that case, we should nevertheless attribute the decision to the group agent and not to a mistaken judgement by the manager if the original commitments were too vague, leaving room for the individual member's judgment to settle the issue.

Shared policies of procedure will not only guide the members' reasoning in the sense that they help them to resolve conflicts. They will also inform them which facts are salient for a decision and, thus, for which facts they will need to gather additional information, including from other parts of the group agent.

If someone fails to participate in those shared policies of giving weight which characterise their role in the group agent, then the group agent's point of view does not guide their deliberation and reasoning on the part of the group. Therefore, their decisions will not count as the group's decisions if it is in conflict with it. More generally, we can say that whether an element c of the point of view of the corporation is relevant for making sense of a certain choice depends on whether there is a shared policy of procedure that gives weight to normative facts of the same kind as c .

One might wonder what remains of the end of group self-intelligibility. What is its role in guiding the member's actions, when they are actually guided by the shared policies of giving weight? I think that there are two different ways to respond. First, the implementation of the shared weight-giving policies will itself be guided by what makes the most sense for the group to do and which way of resolving incompatible ends makes most sense for the group, given its normative profile – its practical identity and previous commitments. Thus, group self-intelligibility will inevitably structure the kind of group its members understand themselves to be a part of. Given that a group's self-intelligibility is constrained by its practical identity, its history and its previous commitments, it is neither free to change completely nor arbitrary. In choosing which shared policies of procedure guide their collective deliberation, for example, giving more weight to environmental concerns in certain situations or more weight to the interests of shareholders in others, the group engage in a process that Velleman has called 'deciding how to decide'. As I have argued at several points in this dissertation, and as we will soon see again, this process is a theoretical endeavour that is not

completely up to the members.

Shared policies of procedure and shared policies of giving weight make the process of being guided by group self-intelligibility an essentially shared and social endeavour. It is not something that the members can do individually. Rather, it is embedded in a social structure that is both formal – that is, structured by the organisational structure – but also substantive given the shared policies of giving weight that help members resolve conflicts and pick out those facts that are salient from the group's point of view for the decision they have to make for the group.

5.5 Self-Intelligibility and a Group's Practical Identity

In the previous section, I argued that shared policies of giving weight undergird our practical identity, thereby providing the group with a mechanism to influence its members' deliberation in relevant ways. The practical identities of individuals also help us to distinguish between those ends that give rise to autonomous actions and the ends that are external to the agent and undermine its self-governance. To serve these functions, the ends must be stable across time. This stability is also required for self-governance across time: diachronic self-governance.

5.5.1 A Diachronic Argument

The practical identity of a group is a source of self-intelligibility and, thus, of practical reasons. This means that being able to change one's practical identity at a whim entails an arbitrary shift in the reasons that apply to the agent.³² If this were, indeed, what the normative landscape for groups looked like, we would find it difficult to have long term relationships with them, since those relationships depend on them sticking to their word or, more generally, entering into agreements and mutual assurances in which one does not reconsider one's commitment to the agreement on the first sight of short term benefits or other temptations. All of these types of actions – agreements, persisting resolutions, but also threats – depend on an agent's ability to rationally hold persisting intentions. In a recent paper, I have argued that the rational requirement which governs the diachronic rational stability of intentions presupposes the continuity of the agent's practical

³²This argument is discussed in Velleman (2006, Chapter 12). It also appears in Bratman (2018) under the heading of 'fragmented reason'.

identity.³³ There, I argued for a counterfactual rational requirement in the spirit of David Gauthier's account of diachronic rationality.³⁴ The central idea is that the rational stability of an intention depends on whether or not it is rational to reconsider the previously formed intention in light of new information. This depends on whether the information presently available to the agent would have changed the normative assessment which was arrived at in previous deliberation and resulted in a corresponding different decision, had it been anticipated at the time of intention formation. If this is the case, then it is rational to reconsider the intention now. Thus, the rough idea is that information must be relevant to that sum of one's previous reasons which gave rise to the intention in the first place.

The stable practical identity plays a role if the decision is based, in some sense, on the constitutive ends, values and deep commitments which make up the practical identity. A change in one's practical identity, I argued, typically cannot be easily anticipated in deliberation. This means that once the practical identity changes it will always change the previous normative assessment that gave rise to the intention, thereby making it rational to reconsider the intention. Thus, a changing and easily-alterable practical identity will undermine or even cancel the rational stability of an intention, especially in cases of long-lasting intentions, and make it rational to reconsider previously formed intentions that were part of an assurance, a threat or a resolution to resolve a conflict between incommensurable ends. Therefore, to have these types of actions open, e.g. assurances, threats and resolutions, agents need stable practical identities. These types of actions are not only in the agent's self-interest in that they are also necessary if one wants to interact with other agents. An easily changeable practical identity diminishes these abilities and the agent's potential for participating in interpersonal relationships with other agents which rely on the possibility of giving assurances to which the agent can commit.

This argument rests on the merits of my particular account of diachronic rationality. I will not reiterate those arguments here beyond their gist, which I have already touched on earlier. The account of diachronic rationality explains how an act of will can resolve conflicts among incommensurable choices, as well as explain how and when it is rational to refrain or not refrain from giving in to temptation, all while remaining true to the typical intuitions voiced by Holton (2004), Bratman (2012) and Van Fraassen (1973). In doing that, the view provides us with an account of how we can exercise

³³See Altner (2021, section 4.4).

³⁴Gauthier (1994, 1998)

autonomy over our future actions.³⁵ The fact that other accounts, such as those of Ferrero (2010, 2014) and Broome (2013), fail to capture our different intuitions in these cases speaks in favour of my conception of diachronic rationality. Nonetheless, other accounts, like that of Bratman (2018), might provide us with a similar argument that also highlights the importance of a stable practical identity for diachronic self-governance and autonomy.

The Group's Practical Identity as Externally Grounded

How does a group's practical identity achieve the kind of stability required for diachronic self-governance? It might be due, in part, to the stability of the individual's contributing mental states. Perhaps, as an empirical fact, it just happens to be the case that individuals represent the group in the same way throughout their period of membership. But this wouldn't explain how, as we often see in groups, the group could retain its stable practical identity over and above the lives of its members. In general, the group is always vulnerable to members occupying influential roles within a group agent that allow the corporate body to be exploited for their own benefit which, again, provides individual members with incentives to represent the group in ways coherent with their desire for self-enrichment. This is what we saw in the misappropriation case. Moreover, a group's internal policies, whose stability depends on being shared among several members, can also be liable to change in similar ways. All in all, none of the elements that constitute a group's internal social architecture seem to be able to ensure sufficient stability for a group to be capable of diachronic self-governance. If this is the case, there is no compelling evidence encouraging outside agents interacting with it to engage in agreements with the group, given that they cannot expect it to have diachronic self-governance reasons to remain committed to these agreements. In my view, the only way by which a group may achieve a sufficiently stable practical identity is by grounding its practical identity externally.³⁶

In what follows, I will propose that the practical identity understood as

³⁵This is how Velleman (2009b, Chapter 9) characterises Gauthier's aim in his paper 'Assure and Threaten', but it also captures my aim in 'A counterfactual account of diachronic rationality', albeit, there, I explain what I do as giving an account of the diachronic functions of a will.

³⁶In L&P's account of self-identification, the forces that shape a group agent's practical identity are the emotions members feel when they are invested in the group. Feeling shame and pride on behalf of the group leads them to attempt to sculpt the group's character via owing and disowning actions and parts of the group's history. Thus, in their account, the group's practical identity is not externally restrained. This, too, leaves the practical identity of the group too fluid – as highlighted by the misappropriation case – and unable to distinguish between intentional actions by the group and mere group activities.

externally grounded principles of choice comes in the form of social practices that define the role of the corporation within society, but also roles within the corporation, such as the role of the CEO. Since the content of these social practices is not up to any one corporation or the members of corporations, it cannot be changed unilaterally. They are stable across time and shape the deliberation of the members. They help to distinguish between the aims that can be attributed to the corporation and the aims that are external to it, undermining its self-governance both at a time and across time. The kind of social practice that can play these roles has been studied under the name of *constitutive rules*. The idea of constitutive rules was first described by Rawls (1955), explored more thoroughly by Searle et al. (1995), Searle (2006) and applied to the context of group agents by Hindriks (2003, 2008) and Ludwig (2017b).

Constitutive Rules

The notion of the constitutive rule was first popularized by Rawls' seminal paper 'Two concepts of rules'.³⁷ There, he distinguishes between rules that are the summary of like decisions under similar circumstances, in which the decision comes prior to the rule, and rules that constitute an activity and that are logically prior to the activity they govern. The latter are constitutive rules. This insight was subsequently explored and developed more fully by Searle (2005, 2006) and Hindriks (2003), among others.³⁸ According to Searle, constitutive rules play a central role in our social world: institutions such as money, marriage or a university may be understood as systems of constitutive rules. Constitutive rules create social facts and, thereby, make up the social world. These rules build on the idea of a status function. A status function captures the collectively accepted function or purpose that is imposed on some entity. This happens, for example, when we collectively agree to treat some set of pieces arranged on an eight by eight square board as a chess set. A status function has the following formal structure: 'X counts as Y, under conditions C.' Since, in our example, chess pieces can take infinitely many physical forms, the function performed by the various pieces can't be in virtue of their physical features. Thus, it is necessary for a status function to be collectively accepted. Collective acceptance can be understood through the concept of shared intentionality, developed in different ways by Gilbert (1990, 1992), Tuomela (1995), Bratman (1993) and Searle (1990). Here I adopt, as I have elsewhere in the dissertation, Bratman's account of shared intentions (Section 2.4.1). Since an object can be more or

³⁷Rawls (1955)

³⁸See also Ludwig (2017b).

less good at fulfilling a function, by assigning a new function to an object, a status function introduces a teleological normativity.³⁹ To fulfil its function, a status rule also comes with a deontic status consisting of characteristic rights and duties. For example, the remains of a wall might be collectively accepted to count as a border, and that acceptance comes with the obligation not to step over it and the right to use the land on your side of it as you see fit. Thus, the deontic status will include a number of rights and duties. To make this explicit in the formal structure of the rule, (Hindriks & Guala 2015, 474) make the following change:

Definition 5.5.1. *The Constitutive Rule:*

If circumstances C hold, then X counts as Y , and if Y , then Z .

The variable Z now stands for the content of the imposed institutional term Y , whose content refers to the set of rights and obligations of which their deontic status consists. Consider the following example. Take C to be a chess board with the pieces assembled on their starting squares. Then, the piece on square E1 counts as the white king. If this piece is the white king, then the player who plays white is *allowed* to move it at most one square if that square is free and has a duty to recognise her loss if the king can't evade capture.

Note that the assignment of a status function involves special cases of two of the basic building blocks of social reality: the assignment of function (the function of 'counts as' representation) and collective intentionality. The collective acceptance of a certain function is constitutive of the deontic status and the deontic powers that come with such a status, like the rights that a border confers to the one whose land thereby gains the status of property. Although not made explicit in Hindriks (2003), the deontic status enables the realisation of the function. In a later paper with Guala, Hindriks says as much:

[...] deontic powers contribute to the ways in which institutions actually function. [...] [I]nstitutions can be explicated in terms of normative rules. Such rules feature deontic powers. And partly because of that, those rules promote cooperation. This, in turn, is conducive to the performance of institutional functions. The rules feature in the attitudes of the participants of an institution, perhaps as normative expectations.. And they typically affect the preferences that agents have such that they will be inclined

³⁹See Hindriks (2003, section 2).

to cooperate or do so more robustly. In this way, they have a causal impact and contribute to the persistence of institutions.⁴⁰

Thus, the deontic status helps to ensure the institution's etiological function, because it explains why the function of an institution exists and persists.

A status function becomes a constitutive rule, and, thereby, gains an additional kind of normativity, when it becomes a matter of general policy.⁴¹ Searle explains the idea of a general policy by appealing to the notion of codification. Minimally, codification refers to a rule being formalised and written down in law. The normativity involved is twofold. *Codification* creates the possibility of abuse or of rule violation, while also creating a certain kind of *self-referentiality*: a constitutive rule only exists because people accept collectively that it exists, making it possible for there to be genuine tokens of the type related to the rule. 'About particular tokens [of money],' Searle argues, 'it is possible for people to be systematically mistaken. But where the type of thing is concerned, the belief that the type is a type of money is constitutive of its being money'⁴²

Constitutive rules create the social reality by creating social facts and social practices. They make possible practices like playing chess that wouldn't otherwise exist. These social facts given rise to by constitutive rules cannot be changed unilaterally by a subset of agents who participate in them. Rather, they are true in virtue of enough people believing and participating in them.

5.5.2 Constitutive Rules and Practical Group Identities

The idea that constitutive rules are shared representations allows us to say that they feature as representations of normative rules in the point of view of the group. The concept of constitutive rules fits neatly with the overall cognitivist picture that we have adopted from Velleman, through which we construe group agency through the constitutive aim of self-intelligibility.⁴³

This brings me to the following necessary condition on the practical identity of structured groups. Let O be any generic practical identity of a group G , stated as a principle of choice such as a for-profit corporation's aim of maximizing shareholder value. Further, let P be a type of group and let C

⁴⁰See Hindriks & Guala (2019, 15).

⁴¹See Searle et al. (1995, 48).

⁴²Searle et al. (1995, 33)

⁴³The difference between the constitutive strategy with regard to agency is that the conditions for someone to count as an agent aren't up to us. We can't change these conditions like we can change the rules in a game of chess or football, for example.

be any set containing the conditions for setting up constitutive rules relevant to different group agents, such as a certificate of organisation that is signed by the appropriate authorities, along with certain kinds of organisational structures and so on.

Definition 5.5.2. *External Practical Group Identities:*

*If a constitutive rule R – of the form ‘ X is represented as P under conditions C ’ – applies to group G because of the presence of a sufficiently ubiquitous social practice whose content is R , **then** the normative status O that is implied by P determines part of G ’s full practical identity.*

The definition is pretty straightforward. It equates the deontic status of an effective constitutive rule with parts of a group’s practical identity. It is necessary insofar as it is required for robust (diachronic self-governance).

The advantage to this definition is that we can now answer certain questions in a more structured and pointed way and we can make use of the findings regarding the properties of constitutive rules:

1. What puts a social practice into place and how does the practice become part of the point of view of the group agent?
2. Does the external practical group identity, conceived of in this way, play the roles that we have identified earlier?

Anchoring a Group’s Practical Identity

Brian Epstein argues that questions about what puts a rule into place can be understood as questions about what anchors the rule in our social world.⁴⁴ In the case of constitutive rules, this question is easy to answer. The rules are anchored by those facts that make people join and participate in the shared intention of representing a group agent as a type of agent like a public corporation, a non-governmental organisation, a religious institution, a sports club, a court or the state as a whole. Although this is true, it can’t be all there is to it. To see this, consider the case of the public corporation. There, the deontic status that is implied by the constitutive rule does not only consist of the rights – such as the right to sue or to borrow money – but also of the duties that are transferred to the organisation by the power of a constitutive rule. The duties that apply to the public corporation are a matter of contention in the fields of legal and economic theory, as well as business ethics. Some argue that the public corporation in particular, compared to other corporate forms, should be governed in its shareholders’ interests, while others maintain that

⁴⁴See Epstein (2015, 2019).

the duty of any corporation is to balance the interests of those parties who have a stake in it. The fact that one can represent the public corporation in two ways means that there are at least two possible constitutive rules.

My claim is that whether the corporation is represented as having the duty to maximise shareholder value or to balance stakeholders' interests is a *theoretical question*. Epstein calls this the 'grounding inquiry'. It asks what principle explains a certain phenomenon. It is a question about what social facts obtain or, put differently, which social practice prevails. Arguments put forward by defenders of both of these potential principles of choice play a role in that they can function as the reason and explanation for a person's participation in a specific practice. Importantly, practices may differ from one social or market environment to another. In competitive markets, shareholder governance might be the salient practice because it represents the dominant strategy over other corporate governance strategies. In less competitive markets, the totality of reasons might speak for the practice of stakeholder governance. Furthermore, different legal requirements on the organisational structure of public corporations, like governance by a board structure or participation by employee representatives, may impact the deontic status which applies to the corporation. These questions will be discussed in detail in Chapter 6.

These practices anchor the practical identity externally because the members of the corporation are not their only participants. Other participants include the group's shareholders, academics (the orthodox interpretation of the corporation in microeconomics courses remains that the firm should maximise profits), the media, legal experts and politicians. If enough people within the relevant environment represent the public corporation as one which ought to, say, maximise shareholder value, then this participation anchors that very same status. In a certain sense, I have just repeated the same claim, but have not justified how the principle of choice contained in the constitutive rule becomes part of the point of view of the group. Let me offer two brief arguments.

The first argument is that constitutive rules so conceived explain the business judgement rule from Section 3.2. That rule states that if a choice by some agent on behalf of the corporation cannot be rationalised by reference to the aim of benefiting the corporation as an independent entity, the member must bear responsibility for the action individually.

The second argument concerns similar cases from other practices. Suppose that you are playing chess, but, realizing that you will lose unless your king makes an illegal move, you move the king as if it were a knight. Even if your partner would accept this as a legal move – maybe because she has pity with you – the game would no longer count as a valid chess game at a chess

competition, and neither one of you would count as having won the game. Or, suppose that you have won by relying on the help of a chess computer. It is partially constitutive of playing the game that you choose your moves without the help of a computer. But, in this case, your game would count not as your game, but as one played by the computer and, at a competition, you would similarly be punished if not excluded from future matches altogether. These phenomena only make sense if the constitutive rule goes beyond your own agency, which is part of the very definition of constitutive rules as shared intentions in which you participate. If these rules can determine whether some activity counts as your action or not depending on your own mental states, it clearly follows that, in the case of corporations, constitutive rules can similarly explain the existence of certain mental states even if the agent doesn't itself endorse the end in question.⁴⁵

This fact is only surprising if we have a conception of agency which postulates some tight metaphysical connection between the agent and the body. But groups are social entities, the very possibility of which presupposes the existence of larger social structures within which the agent is embedded and without which they couldn't fulfil the specific functions required of them. So, it is unsurprising that this larger sociality seeps into some of the agential structures of group agents like corporations.

We now have at our disposal the tools to solve the interesting case in which the members of the group agent do not represent the group as of type *P*, despite the existence of such a social practice. This represents a failure of theoretical rationality, similar to a case in which one fails to treat a genuine dollar bill as a dollar bill.⁴⁶ Or, consider a case where someone presents some *X*, as *Y* under conditions *C*, in accordance with the relevant constitutive rule, but fails to correctly interpret the deontic status *Z*, which is the bundle of rights, privileges and duties that come with this presentation. In that case, one believes that one has a genuine dollar bill, but does not believe that it carries the agreed-upon value and isn't prepared to exchange the relevant goods for the bill. Returning to the corporation, one might recognise that the relevant social practice comes with a deontic status of maximising shareholder value, but not believe there to be compelling reasons to join the practice. Then, failing to rationalise one's actions in light of the deontic status will mean that one fails by the aim of self-intelligibility, meaning that one's choice on behalf of the corporation does not achieve the status of a full fledged corporate action.⁴⁷

⁴⁵This argument will be picked up in Chapter 6.

⁴⁶This also captures the kind of rational failure that Williams (1981, 101) wanted to exclude from his conception of reasons, e.g. reason statements that rested on a false belief.

⁴⁷The overall explanatory strategy in this section is in the spirit of what Haslanger

The Different Roles of the External Practical Group Identity

Once a social practice is established that plays a role in an external practical group identity, it can play three characteristic roles.⁴⁸ First, it shapes a group's *private self-conception* of itself and, thus, how the members represent the group to themselves. It does so by giving them reasons to integrate the practical identity through shared policies of giving weight that assign the practical identity a coherent weight relative to other normative considerations. This provides further derivative reasons for internal coherence and unification in the sense that Rovane (1997) has argued for. Thereby, it makes diachronic self-governance possible.

Second, a social practice also provides the group members with reasons to represent themselves as a group of type *P* to others, since being of a certain type comes with advantageous rights and claims. Thus, it influences the group's *public self-description*.

Third, for agents outside the group who participate in the constitutive rule, the group *G* counts as *P* if they believe the relevant conditions *C* are satisfied. This will lead them to come to a *public description* of *G* in terms of *P* which comes with relevant expectations and claims *O* typical of the normative status. Thus, the practical identity will influence how other agents interact with *G*.

Further notes on Group Practical Identities

To conclude this section, I will highlight certain other facts that can anchor practical identities. These might be facts about the organisational structure which are also contained in the constitutive rule, but which are specified within the conditions rather than the status of the rule. For example, consider management by a board structure. Such a condition can have its own normative consequences without being contained in the deontic status.⁴⁹

Other facts that anchor the group's practical identity supervene on the members' skills, abilities and personal ends which partially anchor and ground the group's abilities and its own ends. A corporation that drills for oil will have acquired skilled members whose main if not only qualifications might lie in this area and whose idiosyncratic competences and interests will shape the group's own interests. These members will rethink their membership should the group shift its material focus to areas beyond their speciality. Such ex-

(2016) has called social structural explanations that explain the behaviour of an entity by appealing to the structures of the system it is a part of.

⁴⁸Thanks to Grace Paterson for pointing this out.

⁴⁹See Section 6.3.1.

planations will play a more important role in informal groups which exist within looser frameworks. An informal book club, for example, is mostly bound by properties internal to it, like its members' personal interests and abilities, and could in principle very easily turn into a group that protests for the right to free speech if the associated ends coincide with the members' interests. In this case, there is still something of a practical identity, although this is obviously not external in the sense above. This also means that there might be other strategies which could explain the possibility of synchronic and diachronic self-governance by the group. I have not pursued these here, but also do not believe that these alone could explain how group's are robustly able to manifest a will.

Similarly, the group's own history will partially constitute what it makes sense for a group to do, but also which values to adopt. Here, again, L&P's more affective story, which involves investing in a group agent and feeling shame and pride on behalf of the group's achievements, can play an important role in explaining how the group's history may influence what currently makes sense to it.

Finally, the group's own plans and commitments, embedded within shared policies of giving weight, will sculpt the group's practical identity. Part of the normative force of these commitments will stem, transitively, from how they connect to the group's external practical identity.

5.6 Responding to the two Challenges

5.6.1 The Attributionist Challenge

It is easy to see how the constitutive account provides us with the right answers in the misappropriation case. Recall that in the misappropriation case, actions that lead to the demise, harm or decreased well-being of the corporation do not make sense from that point of view which includes either the aim of increasing shareholder value or balancing stakeholder interests. Both of these aims, it has been argued, capture the well-being of the overall corporate enterprise. The former does so in that shareholder value is just the overall value of the corporation. The latter aim is threatened insofar as a corporation's demise would leave all of the stakeholders' interests unfulfilled.

5.6.2 Self-Intelligibility and Responsibility

The ODV is easily translated into the constitutive account of group agency. It is a group with a point of view shaped by a practical identity that consists of

a single constitutive, non-moral, maximizing aim, such that this aim shapes the reasoning of the group by being both its private self-conception and its public self-description, thereby giving rise to a web of shared policies of giving weight that are characteristic of the ODV.

The issue now comes down to the question: Is the ODV capable of self-intelligible actions and, thus, capable of acting autonomously?

In one sense, there is an affirmative answer to that question. Members may choose on behalf of the ODV in light of its mental states, beliefs about the circumstance and its practical identity. Aiming at making sense from the group's point of view will shape how they interpret and 'act out' a specific choice by the group. The aim of self-intelligibility will also lead them to intend the necessary means of existing ends and intentions and to resolve conflicts and inconsistencies among its commitments, thereby ensuring that the rational requirements are satisfied as a means of choosing in a way that makes sense from the group's perspective.

Now, suppose that some group's constitutive, practical identity defining aim *A* is incompatible, on a regular basis, with the demands of morality: so that morality demands that some obligation *O* is satisfied, where *O* is inconsistent with the necessary means *M* implied by its constitutive end. If this is the case, then it is instrumentally irrational for the relevant group members not to *M* and, thus, it is not self-intelligible to do so. If the group members were nevertheless to decide on the group's behalf to *O*, then the group wouldn't act in a fully intentional way. Having *A* as their constitutive end, the group agent has manoeuvred itself into a blind alley, morally speaking. This at least is how Velleman gauges the situation of the 'immoralist'.

Shedding his identity as a mobster would be a betrayal of the mob and hence of the commitments fundamental to that identity. Hence it is not a rational step for the agent to take, all things considered. But the act of reconstituting his set of reasons is indeed supported by a crucial subset thereof - namely, the reasons arising from his underlying identity as a rational human being. And the agent can act on that subset of reasons while holding the others in abeyance; for he can think of himself merely as a human being, reflecting with critical detachment on his more specific identities. Thus, he can tentatively suspend his identity as a mobster for the sake of considering whether to reject it altogether.

Even this tentative suspension of an identity would not be rational for the agent, all things considered. His commitments to the mob strongly militate against even toying with the idea of betrayal. But he can still toy with the idea, albeit irrationally.

Indeed, he can literally toy with it, by playing or pretending for a moment that he isn't committed to the mob. He can imagine himself to be only a part or aspect of everything that he is, so as to make believe that he is deciding from scratch what to be.⁵⁰

The problem for the ODV is that, in the face of systematic immorality, there will be no crucial underlying identity as a rational human being to fall back onto. In the end, groups like the ODV are tools set up for a specific purpose and, if the set-up and the constitutive purpose are immoral, then they, along with their members, are stuck with that purpose. On the other hand, the members can, in most cases, fall back to their identity as a human being and climb down from the rotten social branch that is their group. For states, this procedure might be difficult: in such cases, it makes little sense to reform an amoral agent such as the Nazi regime or the mafia. Instead, it is more productive to create an alternative agent, whose practical identity is sensitive to the range of moral facts and whose constitutive aim does not force it into being, not an immoral, but amoral agent. One that cannot participate in our moral practice of responsibility.

So, let us return to the question of whether the ODV can act fully autonomously. Being a free author of one's actions means inventing one's future conduct. The restrictions posited by the ODV's practical identity at least restrict the range of intelligible actions to those not incompatible with its practical identity. As stated above, this does not mean that it cannot choose how to implement its constitutive ends. Rather, a certain kind of agency – namely, moral agency – remains out of reach, if it is systematically at odds with its constitutive aim. If part of the goal of self-intelligibility is choosing one's own ends and abandoning them when required, then the actions of the ODV cannot live up to the constitutive standard of agency. This seems to be at least implied by the following quote from Velleman.

"the feeling of metaphysical freedom, though illusory, does tend to track a real class of actions - namely, the actions in respect to which the agent is epistemically free. And I say that we have a legitimate interest in identifying actions belonging to that class, because they are the actions for which there will be something to gain from subjecting the agent to praise or blame, reward or punishment.⁵¹

The ODV seems to lack this kind of feeling of metaphysical freedom. Having one's set of options restricted in a way that bars responding to moral

⁵⁰Velleman (2006, 311)

⁵¹Velleman (2009b, 71)

considerations, can induce feelings of despair and forlornness which seem to undermine the feeling that one's future is open to oneself. When one has to rank values as the ODV's point of view requires, one relinquishes a certain kind of narrative control on behalf of the group agent. One resigns oneself to a script that one doesn't find convincing, maybe from one's own individual perspective, in which one can only fill in the details. Still, this might allow for some autonomous actions, if the constitutive end is compatible with moral reasons. But if it provides it with non-moral, perhaps solely prudential reasons that outweigh other considerations, then it is not clear whether the group is still able to satisfy the constitutive end of making sense from its own point of view to the fullest extent. And this might be because the individuals, in reverting to their individual perspective, find the group narrative to be internally coherent in itself, but in deep conflict with beliefs held as part of their individual centre of agency.

Or consider another strategy which appeals to the idea of endorsement. Let me briefly quote Velleman:

an agent may understand his behavior in terms of unfortunate traits that do not provide reasons for their behavioral manifestations. Someone who knows himself to be lazy, for example may find his avoidance of work intelligible in that light without thereby finding it supported by reasons. [...] the conception of himself as lazy, rather than as easygoing or laid back, expresses disapproval, which would have to be included in a complete conception of himself. And manifesting laziness while condemning it as such is not altogether intelligible, after all.

. In the case of a corporation whose managers don't endorse their constitutive aim, this aim might nevertheless play its functional role, as it is externally anchored. This functional role will resemble the disposition to be lazy, which shapes the behaviour of Velleman's 'bum'. A possible collective analogue of this lack of endorsement is a case where the corporate members do not participate in the constitutive rule that constitutes the ODV's external practical identity.

These thoughts are rather tentative, but it seems that if we accept that something like the absence of a feeling of epistemic freedom on behalf of the group, or a lack of endorsement, impinges on the possibility of achieving the aim of group self-intelligibility, then the constitutive account tracks the conditions of responsibility and, the powers of reflective self-control. Then, we could conclude that a group with an ODV-normative profile which undermines its members institutional or group's feeling of epistemic freedom

in crafting a narrative on the group's behalf, is incompatible with achieving group self-intelligibility and, thus, group autonomous actions.

If it doesn't track the conditions of responsibility, then the constitutive account seems to only track aretaic-responsibility, that is, responsibility as attributability.⁵² In that case, it would not explain why the ODV should not be included in our practice of responsibility, because it falls short of being capable of a good or an ill will. Then the ODV would, under the constitutive account of group agency, be capable of acting autonomously. However, I hope to have made it plausible that the constitutive account has the resources to capture those imperfections in an ODV's will for which it would be excluded from our practice of responsibility as accountability.

5.6.3 Some Metaphysical Remarks

Although the account of group agency which I have adopted is cognitivist, certain elements to which I have appealed in my account, belong to non-cognitivist theories of rationality. Examples of these are my interpretation of decision mechanisms or aggregation mechanisms and hierarchies as shared policies of procedures, my adoption of the account of shared deliberation and my account of shared intentions, which all stem from Bratman's non-cognitivist theory of action. On the face of it, integrating different aspects from cognitivists and non-cognitivist theories seems metaphysically dubious and might lead to inconsistencies with regard to my constitutive theory of group agency.

Although such inconsistencies are possible, I don't consider them to be a significant danger. Both accounts agree on a wide range of rational requirements that govern which kinds of mental states it is coherent for an agent to possess. Their disagreement concerns the connection between intentions and beliefs and the justification of the norms of rationality. Cognitivists appeal to theoretical reasons in order to explain and justify requirements of rationality. Non-cognitivists, like Bratman and Gauthier, choose an instrumentalist

⁵²Velleman himself does not subscribe to any theory of responsibility. When he mentions responsibility he is quick to point out that his notion of the will as an action guided and shaped by the constitutive aim of agency, is not what is relevant in practices of responsibility. That said I don't think we have to follow Velleman in this regard. His hostile stance is more a matter of his metaphysical commitments elsewhere than what is implied by his theory of action. Fischer (2012) would disagree. According to him, theories that are concerned with self-governance and autonomous actions do not capture all actions for which we want to hold agents responsible. Although I don't agree with Fischer, even if it turns out that he is right, we could still maintain that actions that are guided by the aim of self-intelligibility are more blameworthy, than those in which the agent is not fully in control.

strategy, by pointing out how rationality helps us to get getting what we want in the long run.

Problems might also arise from the fact that Bratman's account of shared intentions rests on his account of individual rationality. He has argued that the norms which govern the rationality of individuals transmit to the collective case. So, his account of shared intentions also holds that a group which shares an intention need not believe that it will execute that said shared intention. This is a difference from the account of Velleman, who argues that intentions constitutively involve a kind of belief that is realized, because of our goal of self-intelligibility, of manifesting our beliefs about what to do. This becomes a problem for those arguments that build on the cognitivist link between group intentions and beliefs, but rely on Bratman's account of shared intentions. I don't think that I appeal to this in addressing the two challenges of the ODV and the misappropriation case, but I might be mistaken. In the latter case, I think that it is easy to simply adopt Velleman's account of shared intentions and to drop Bratman's pragmatic account. Shared policies of procedures and of giving weight are characterized by their content and do not, in explaining the phenomenon of group deliberation and organisational structure, depend on the lack or added cognitive rational requirements.

In this dissertation, I haven't done so mainly, because, from a functional perspective, Bratman's account is worked out to a degree that Velleman's and other accounts of shared intentionality aren't. In Chapter 6, this will help me to straightforwardly explain the impact of specific group identities on the deliberation or self-inquiry of group agents. Self-inquiry, as I argue in section Section 5.4.1, relies on shared intentions. So, I hope that I will be excused for relying on one of the most sophisticated and most well-worked out accounts of shared intention that is currently on the market and promise that, in the face of inconsistencies, I will adopt a purely cognitivist account of agency.

5.7 Situating the Constitutive Account

I have now provided a rough outline of a constitutive account of group agency and explored how it can address misappropriation cases and the problem of the ODV. In the remainder of the this chapter, I will contrast this account with Pettit's theory of group self-identification, Hindriks' Status account of group agency, and Pauer-Studer's constitutive account of group agency.

5.7.1 Self-Identification vs Self-Intelligibility

Philip Pettit's account of self-identification is a rare instance of a non-Davidsonian account of a group will. Self-identification and self-intelligibility share an important property: Both are hierarchical accounts of action. Group self-identification implies an endorsement of attitudes which are either owned or disowned and which may be accompanied by shame and pride felt by those members who are invested in the group agent. It is conspicuous that, no further conditions restrict the kind of higher order attitude. Therefore, it seems that the members of a group, who are invested in it, could own all kinds of group actions. Specifically, actions that also cohere with the members' own individual interests, but not necessarily with the group's interests. For a more extensive discussion of this view, see chapter Chapter 3.

On the other hand, self-intelligibility is, so to speak, the ultimate higher-order goal which is constitutive of agency in the full sense. The will-making relationship in the case of self-intelligibility is not an affective one that involves investing into the group agent, but a cognitive one. The benefits of adopting a cognitive account are particularly evident when they concern groups which are less likely to serve as a cultured medium for affective relationships, such as corporations, but which nevertheless seem to be capable of acting for reasons and of being interpretable via the intentional stance.

5.7.2 Pauer-Studer's Constitutive Account

In 'A constitutive account of group agency' Pauer-Studer develops two objections against L&P's thin account of group agency.⁵³ First, she argues that the account fails to do justice to the normative nature of agency. Part of her objection is that it is not clear how the capacity for reasoning is connected to the functionalist structure by which L&P identify a group agent. Secondly, she points out that L&P's account of group agency does not provide the resources to make sense of the characteristic normative identity of most structured groups, that is, the fact that corporations or, football clubs have different 'characters' and, thereby, different reasons that apply to them. The misappropriation case, in some sense, tries to flesh this out by arguing that, without these normative identities, it is difficult to distinguish between the reasons that motivate the agent to do something and that give rise to purposive actions and those reasons that justify an action.

To solve this problem, Pauer-Studer proposes a constitutive account of group agency which is inspired by Velleman's non-moralizing conception of agency, and the contours of which I have largely adopted in explicating my

⁵³See Section 2.4.

own constitutive account. Pauer-Studer argues that such an account can make sense of the idea that groups act for reasons that reference their normative identity as a corporation, a football club or an NGO. This ‘normative identity, partly made up by the legal form, partly made up by the organizational structure and the specific institutional aims of the group agent, yields the normative standards to assess whether the practices of a group agent allow for a coherent self-understanding.’⁵⁴ The normative identity, together with the aim of self-intelligibility, also plays a role in identifying which decision procedures to choose, which is very similar to the role of group self-intelligibility that I identified with regard to the shared policies of giving weight in Section 5.4.1. Although Pauer-Studer doesn’t make the ontological character of the normative identity explicit, we may conclude that her account is, in many respects, very similar to mine.

The two accounts diverge regarding the relation between the constitutive aim of agency and morality. Pauer-Studer maintains that ‘[i]ntelligibility is a normative requirement driving a group agent towards a behaviour that meets publicly shared ethical standards.’⁵⁵ In Section 5.6.2, I argued that certain group constitutions and, some external practical identities can undermine the fulfilment of actions’ the very standard of success of actions, self-intelligibility. In these cases, the agent’s practical identity makes it unreasonable, from the agent’s current perspective, to act morally. If we were to maintain Pauer-Studer’s insight, then we could still accept the following conditional statement: that a group agent’s constitutive aim drives it towards adopting a moral perspective, if it already has a practical identity, which is, in principle, sensitive to moral considerations. When this is the case, morality and intelligibility could indeed converge, and the latter might drive the agent towards adopting a normative perspective that agrees with the demands of morality.

5.7.3 Hindriks’ Status Account

In his paper ‘A Status Account of Corporate Agents’, Hindriks (2008) provides an account of ‘ordinary collective agency’,⁵⁶ in contrast with L&P’s rationalist conception of group agency. His picture of group agency is in some respects similar to mine. Hindriks, too, suggests that groups like corporations have two ‘faces’. They have an *internal organization perspective*, which consists of a web of jointly accepted procedures for forming collective

⁵⁴Pauer-Studer (2014, 1635)

⁵⁵Pauer-Studer (2014, 1637)

⁵⁶Hindriks (2008, 125).

decisions and, in recent articles, a web of shared policies of giving weights, both of which give rise to a normative group perspective and group actions.⁵⁷

The other face of groups is the *external attribution perspective*. Collective agents like corporations ‘owe their existence to the external ascription or attribution of a status. [...] I call agents of this kind ‘corporate agents’ because a collection of individuals is incorporated into a new entity by being ascribed a status.’⁵⁸ According to Hindriks, corporate agents are distinct from their members, because they have powers that their individual members lack. The individual attitudes of corporate agents, just as in my account, do not supervene only on the members’ attitudes. Instead, the collection of individuals constitutes the corporate agent. The notion of constitution can also be formulated by constitutive rules which carry with them a deontic status. That status, as described in Section 5.5.1, consists of permissions, prohibitions, obligations, as well as powers and abilities.⁵⁹

Hindriks’ writes that both the external and the internal face are complementary. Statuses can influence roles and positions within a corporation, such as the status of a professor or a manager, which rely not only on internal, but also on external recognition. Apart from that, Hindriks’ does not make clear what the relation is between both the two faces. Notably, in recent papers, Hindriks no longer includes the external perspective in his analysis of moral agents. Although in Hindriks & Guala (2019), he has appealed to it in order to explain the functional nature, of groups –, or what as I call it, their instrumental face –, by applying Dennett’s design stance to institutional entities like groups, he does not provide an overarching account how both faces influence each other.

How is the constitutive account different from the status account? In general, the reader will find much convergence. The main difference lies in the theories of action which underlie the respective accounts. In the original paper, the status account seems to imply a Davidsonian theory of action, whereas the recently-added idea of a normative perspective as a necessary requirement for a group will, might be taken to suggest a Frankfurtian, hierarchical account of action. Such an account falls prey to the misappropriation challenge. Admittedly, the availability of the external perspective could be used to counter these issues in a similar way as I have done. But this will probably require a principled account that integrates both perspectives, such as the idea of making sense from a point of view which includes both the inter-

⁵⁷I have also discussed Hindriks’ normative perspective in Section 2.5.5. Overall, his normative perspective is very similar to what I capture under the heading of the sharedness of group self-intelligibility.

⁵⁸(Hindriks 2008, 128–129)

⁵⁹Since statuses must be externally recognised, they are anchored externally.

nal and the external perspective. Otherwise, why should members, from the perspective of the group, include the oughts attributed to the group via constitutive rules in choosing on the group's behalf? Similarly, without a more substantive narrative on how both sides connect, Hindriks' account fails to anticipate social aggregates with the normative profile of a one-dimensional valuer. This is a problem insofar as the ODV fulfils Hindriks' conditions for moral agency. It has a normative perspective, but that perspective does not issue in enkratic beliefs about obligations that are incompatible with its constitutive goal. Given that his account of moral agency also requires relevant member emotions that support the normative perspective, he might have more resources available to exclude the possibility that the ODV is wrongly identified as a moral agent.

5.8 Summing up

In this chapter, I hope to have shown that the challenge to accounts of group wills stemming from misappropriation cases and the challenge to our understanding of collective moral agency coming from cases such as the ODV, can both be met by my constitutive account of agency. The constitutive account provides two new contributions to the debate on group agency. It presents one of the only accounts of a group will and shows why the practical identity of groups must be anchored and grounded externally in order for groups to display synchronic as well as diachronic self-governance. Thereby, it also fulfils one of the goals of my overall analysis, namely making clear the Janus-facedness of structured group agents.

6 Group Self-Intelligibility and Corporate Governance

6.1 Introduction

The identity of group agents is similar to the identity of an individual. Some parts of its practical identity, it can choose itself. A public corporation might, for example, choose what products it produces, while other parts of its practical identity are fixed. Herein lies an important difference between the human and the group agent. Group agents like a public corporation, are designed by other agents so as to have a very specific practical identity.¹ Economists typically don't use the term 'practical identity' when talking about a corporation. They usually say that the public corporation is the kind of entity which *naturally* fulfils a certain function, or at least *should* fulfil a certain function. By this, they mean that the corporation should aim at maximizing profits.

For business ethicists and economists, talk of the function of the public corporation is framed as a question of governance. Should a corporation, especially a public corporation, be governed in the interests of its shareholders? Or, should it be governed in a way that balances the interests of all the parties which have an interest or stake in the corporate enterprise? The former is known as *shareholder governance*, while the latter is known as *stakeholder governance*. I will argue that this debate can be translated into one about social practices and, ultimately, about a corporation's practical identity. Arguing that public corporations should be governed in the interests of the firm's shareholders is typically translated into talk about the CEO's obligations towards shareholders.² This is only one way to spell out

¹Some group agents might not be designed by other agents, but might be able to choose their own practical identity, where this choice is determined solely by their members – for example, a book club or a very tight group of friends.

²In Section 3.2, I discussed how the business judgement rule (BJR) can be interpreted as imposing an end – that can be either interpreted via shareholder governance or stakeholder governance – on the decision-making of the CEO such that decisions must be rationalised

this social practice. In what follows, I will propose two very rough possibilities of spelling it out and show why only one of these ways is reasonable and corresponds to the debates in the business ethics literature.

In the previous chapter, I argued that we should understand the practical identity of groups as constituted by social practices, which in turn, may be spelled out via the notion of constitutive rules. In this chapter, the advantages of this move will become apparent. Using constitutive rules allows us to reconstruct the debate more rigorously, as doing so directs our attention to the necessary conditions for such a social practice to come about, to the important elements that explain how this practice can enter the mind of the public corporation and fix part of its practical identity. For the following, it is important to remember that I am not here taking a side on whether shareholder or stakeholder practice is the better kind of governance. At the end, I will make a claim to the effect that, when shareholder governance is the dominant practice which fixes the external practical identity of a public corporation, this can turn the group agent into a one-dimensional valuer (ODV). As I have previously argued, agents with such a normative profile are not the appropriate target of our reactive attitudes: they should be managed and controlled, rather than participate as one moral agent among many in our political institutions. Anyone who disagrees with this conclusion can, via contraposition, argue that we should, therefore, not allow the external practical identity of corporations to be fixed by the practice of shareholder governance. This conclusion might, by itself, be a reason to reject shareholder governance as describing the right function of corporations.

The chapter will proceed as follows. Since our discussion will revolve around the public corporation and the social practices which governs it, I first specify what a public corporation is (Section 6.2). Then, I analyse the debate on shareholder governance by discussing two potential constitutive rules which might be taken to correspond to different ways of understanding the practice of shareholder governance (Section 6.3.1 - 6.3.2). In particular, these rules differ in the role that shareholder governance is supposed to play with regard to the public corporation and in the reasons why one might adopt or participate in the practice.³ I will argue that one way to articulate shareholder governance is clearly superior and, finally, show how the practice influences the mind of a corporation (Section 6.3.3 - 6.3.4). In the final part

via this end in order to count as the corporation's decisions.

³This follows, in spirit, the distinction that Hindriks & Guala (2019) make between the etiological functions of institutions that are used to categorise different specific institutions and the 'teleological stance' which Hindriks and Guala suggest we should use to evaluate institutions' success at satisfying specific values, from the value of justice or equality to more contentious values such as theological values.

of this chapter I show how one interpretation of shareholder governance has given rise to corporate deliberation and action that is characteristic of the normative profile of the ODV (Section 6.4). I conclude the chapter (Section 6.4.1) by considering whether stakeholder governance is better suited to allow for autonomous corporate actions that are sensitive to moral considerations.

6.2 The public corporation

A group that is incorporated as a public corporation is, according to an influential essay by Kraakman & Hansmann (2017), legally required to include the following features:⁴

- 1) Full legal personality (well-defined authority to bind the firm to contracts and to bind those contracts with assets that are the property of the firm, as distinct from the firm's owners.)
- 2) limited liability for owners and managers,
- 3) delegated management under a board structure,⁵
- 4) Transferable shares.

Kraakman and Hansmann also include a fifth feature in their account of the public corporation which I have deliberately left out: shared ownership by investors of capital. I have omitted it because it is contentious whether shareholders of a public corporation are owners of stock in the company rather than owners of the company itself.⁶

⁴See Kraakman & Hansmann (2017, 1-2), who argue that these features are points of convergence in international law.

⁵Stout & Blair (1999, footnote 9) argue that 'an independent board is one of the most important characteristics distinguishing public corporations from other forms of enterprise. Limited partnerships, LLCs, and closely held corporations all limit investors' liability without requiring them to do business through a board; partnerships, LLCs and private firms provide vehicles for collective investment, sometimes with free transferability of shares; private corporations and some limited partnerships enjoy perpetual existence; and virtually all forms of enterprise permit their owners to delegate the day-to-day management of the firm to hired professionals. In contrast, governance by a board of directors, many of whose members are drawn from outside the firm, seems unique to the publicly held company.'

⁶Ludwig (2017a), for example, follows Kraakman and simply states that this is the case. Kraakman seems to believe this to be a matter of some kind of logical necessity. Later on (Section 6.3.2), I will discuss some of the arguments for why it is at best odd to assume that shareholders own or should own the corporation in any meaningful sense of the word.

6.3 The Practice of Shareholder Governance

Analysing the practice of shareholder governance via constitutive rules has two advantages. First, constitutive rules involve shared intentions in the sense of shared acceptance of the counting-as-representation. I have adopted the account of shared intentions offered by Bratman, who argues that participating in a shared intention requires that you have reasons for doing so such that these reasons persist in a sufficiently interrelated manner with the reasons of the other participants in the shared intention. Focusing on shared intentions allows us to structure and better understand the business ethics debate about shareholder and stakeholder governance. Specifically, this focus helps us to lay out the reasons why managers should participate in one of the two practices.

Second, in involving the assignment of function, constitutive rules help us to discern how the function of maximizing shareholders' interests comes into play. As some have suggested, it might simply be a way to *categorise* the particular kind of cooperative enterprise which is involved in large-scale productive activities for which shareholders supply the capital. Or, it might be tied to characteristic *obligations* that underlie the relations between participants in a firm, where these relations are essential or characteristic of the productive enterprise that the public corporation participates in. How the function features in the constitutive rule will shape the kind of role that it plays in the resulting social practice.

The discussion to follow will have the following structure. In both sections, I will first spell out the constitutive rules that I take to characterise the social practices that make up shareholder governance, analyse why and where the function of profit maximisation comes into play and what its resulting role is. Next, I will present the arguments that have been forwarded in the business ethics literature for this particular way of assigning the function to the public corporation. It will turn out that only the second way of understanding shareholder governance as a social practice corresponds to how shareholder governance is actually understood in the business ethics debate on corporate governance.

6.3.1 Essentialist Shareholder Governance

Consider the following constitutive rule.

Definition 6.3.1. *Essential Shareholder Governance (ESG-Rule):* *If an organisation X satisfies the legal conditions highlighted in Section 6.2 and*

the board of directors aims at maximising its shareholders' interests, then X counts as a public corporation.

First, note that, given our previous analysis of constitutive rules, the ESG rule's formal structure is incomplete. If it were spelled out fully, the rule would have an additional conditional at the end to specify the deontic powers which accompany the status of a public corporation, such as being able to sue people or other corporations, being able to take on debt, and so on. The structure is also incomplete insofar as I have, for the sake of simplicity, specified that it need only to satisfy the legal conditions highlighted in Section 6.2, although in practice there will be many more specific requirements that it must fulfil, like being registered at the state where the corporation conducts its business (in the US), not having filed for bankruptcy, and so on. What distinguishes this rule from the one to follow is that the aim of maximising shareholders' interests (which, as we will see, is typically taken to come down to the end of maximising profits) features in the conditions for the assignment of the status of a public corporation. It says that for the assignment of the status and the associated rights – suing others, taking on debts, and so on – making profits is one of the conditions that establishes this assignment. It explains why the organisation *X* is a public corporation.⁷ To find out whether the constitutive rule is true, we must ask ourselves whether the different conditionals in the antecedent are all relevant or required for the 'count-as' relation to be put in place. Does having the aim and satisfying the legal conditions suffice to make us count something as a public corporation? Put differently, is having the aim of maximising profits relevant for us to count something as a public corporation, or is it, as I will argue later on, just a consequence of something already counting as a public corporation? If it turns out that we need the aim in order to count something as a public corporation, then the ESG-rule is the correct way to describe the practice of shareholder management. If not, we must look elsewhere.

Goodpaster (1991, 55), an influential voice in the debate on shareholder versus stakeholder governance, might be understood to argue for the ESG-rule in the following passage. In it, he speaks about the kind of group a corporation turns into when its managers aim to balance its stakeholders' interests.

For it blurs traditional goals in terms of entrepreneurial risk-taking, pushes decision-making towards paralysis because of the dilemmas posed by divided loyalties and, in the final analysis,

⁷In the terminology of Epstein (2015), it metaphysically explains or grounds the social fact of being a public corporation.

represents nothing less than the conversion of the modern private corporation into a public institution and probably calls for a corresponding restructuring of corporate governance (e.g., representatives of each stakeholder group on the board of directors).

Goodpaster believes that the corporation turns into a different kind of organisation, namely a public institution, if its managers would try to maximise or further the interests of different stakeholder groups equally without prioritising the interests of shareholders. Thus, he sees it as constitutive of a public corporation to have the primary aim of profit maximisation. Boatright echoes Goodpaster, claiming that ‘there is some value to maintaining the corporation as a private sector institution with a strong profit-maximizing imperative’,⁸ which means that, without the aim of profit maximisation, the public corporation would not be the same kind of institution.

I take these authors to suggest that the aim of profit maximisation is a necessary condition for a public corporation to count as one in the first place and, thus, to argue for a practice of public corporation governance which may be spelled out by the ESG-rule. By playing this role, the end of profit maximisation is used to categorise the organisation by way of its role in solving a problem of efficiency and welfare allocation. Only a practice that ties the aim of profit maximisation to the nature of a public corporation allows the profit maximising corporation – a necessary element in the free market – to provide the kind of social utility that it does. Now, we will venture onto a little excursion: the argument of efficient welfare generation and allocation in which a shareholder-governed firm plays a key role. Afterwards, we will touch on another efficiency-related reason why shareholder governance is supposed to be the best solution to solve the problem that certain kinds of firms face when producing their goods and services. Both of these arguments take the aim of maximising shareholders’ interests to be a defining feature of the public corporation. Both try to give arguments supporting this aim and, thus, argue that managers of public corporations should participate in the ESG-practice.

Shareholder Governance and the Welfare argument

The basic idea behind why managers should act in the interests of shareholders in order to maximise societal welfare in a Pareto optimal way has been nicely summed up by Jensen (2001, 11). Social welfare, Jensen argues,

is maximized when all firms in an economy attempt to maximize their own total firm value. The intuition behind this criterion

⁸Boatright (1994, 401)

is simple: that value is created [...] whenever a firm produces an output, or set of outputs, that is valued by its customers at more than the value of the inputs it consumes (as valued by their suppliers) in the production of the outputs. Firm value is simply the long-term market value of this expected stream of benefits.

Under certain heavy idealizing assumptions, the idea that all firms in an economy attempt to maximise their own firm's value by maximising profits has two important effects on society's overall welfare.

First, welfare is increased by the amount of profit that each company makes, since the amount of resources it takes out of the economy – the costs of its products – are balanced by the price that consumers are willing to pay for them, which, given the company's profitability, is greater than the input price. The idea here is to not only to think of current profits, but of the stream of profits over time, since some products (like the energy produced by a hydropower plant) might require greater investments at the beginning and only becomes feasible over time.⁹ The expected stream of profits is discounted by individuals in the stock market and captured by the firm's overall market value. This brings us to the first important assumption which the argument requires. For the firm's overall market value to be a sensible indicator of its social value, all of the costs entailed by the production of the output must be borne by the firm and included in the price. In other words, there must be *no external costs*. Economists have argued that this can be addressed either through legal regulation or by assigning to some private entity alienable property rights over all the relevant factors in the production and consumption of the goods and services.¹⁰ A firm that aims to maximise its market value in this way – which, as I have already stated, is intimately connected to maximising long-term profits – will do so by expanding its supply as long as each 1 € of input yields at least $1 + x$ € of income. In a perfectly competitive market, the profit x will be 0, since any firm would have an incentive to undercut the other firm's price so long as that price remains greater than the costs of production. If any one firm were to fail to reduce prices, it would lose its market share and fail to sell its products. If, on the other hand, a company has a monopoly, it can charge a price which far exceeds its due to the absence of competition which could undercut its prices. Thus, the argument goes, for-profit firms are the most socially-efficient form of organised production in a market without monopolies and no external

⁹See Jensen (2001, 11 ff). An in-depth discussion of this argument is presented by Hussain (2012).

¹⁰See Coase (1960), the seminal paper on externalities for the solution of assigning property rights to counter external costs.

costs.

Second, when a firm makes a profit, it will partially forward that profit as dividends to its shareholders. The reason why *all* firms in an economy must aim at profit maximising can be explained as follows. Suppose that rational investors invest in firms to make money. Assuming efficient markets, rational agents will invest in firms or create them only if, by doing so, the marginal gains or profits are higher than they would be if the agents were to invest their money in other firms or produce goods themselves. Thus, they will invest in those firms which maximise the expected value of their investment. The allocation of resources done in this way will be Pareto optimal, which means that there is no other distribution of resources available in which no one is worse off and at least some are better off.¹¹ If certain managers were to not maximise profits, this would distort the allocation of resources. Presumably, there would be a better way to invest the money for the rational investor that would make everyone better off. Therefore, managers should act in a shareholder-orientated way, so as to ensure the Pareto optimality of investment.

Although the premisses that the consequentialist argument demands – no external costs and no monopolies – are seldom met in practice, according to Jensen (2001, 16), this does not undermine the necessity to conduct business in a way that maximises shareholder value. He claims that

resolving externality and monopoly problems is the legitimate domain of the government in its rule-setting function. Those who care about resolving monopoly and externality issues will not succeed if they look to corporations to resolve these issues voluntarily. Companies that try to do so either will be eliminated by competitors who choose not to be so civic minded, or will survive only by consuming their economic rents in this manner.

It is unclear Why the economic rents of corporations that cause huge external costs are still desirable, on the level of overall social welfare. Thus, Jensen's comment confuses the normative argument – why corporations *ought to* make money – with a positive argument which shows why profit-making is a necessary condition for a corporation to exist and persist, thereby undermining the normative strength of the former. A more appropriate response to the problem of external costs has recently been development by Heath (2014), who argues that corporate responsibility consists in voluntarily resolving external costs by the corporation itself, which leaves the welfare argument unscathed

¹¹See Quinn & Jones (1995) and Sen (1985) for explanations of why profit maximisation in a perfect market leads to a Pareto optimal solution in the overall economy.

by protecting the efficiency function of the corporation in an imperfect market.

Here is why one might think that Jensen's argument implies the ESG-practice. Without corporations aiming effectively at profit-maximisation, the market, defined as an efficient system of resource allocation which satisfies the consequentialist argument, would not have the property of being a system for the efficient and fair allocation of resources. For the market to satisfy the consequentialist argument, it is necessary that corporations, as a matter of fact, aim at maximising profits. The public corporation characterised by the ESG-practice is, thus, a constitutive element in the overall argument. Without public corporations which aim at profit maximisation, the market that requires these agents to be there, for it to be a means to efficiently create welfare, would be a different institution, it would not be a free market. The firm inhabiting such a system would no longer be a public institution. It would be a state-owned firm or a non-profit organisation instead of a private sector institution that satisfies different values. But there is the rub, for our ESG-practice is concerned with the status of a public corporation and not whether that corporation is a private sector institution. We can easily think of stakeholder-governed public corporations occupying a free market, if the competitive pressure does not require them to maximise profits. So what the welfare argument comes down to is an argument for, at best, a *good* public corporation. The public corporation which the argument supports is good because, under the heavily idealised premisses assumed in the consequentialist argument, it increases the overall welfare of society in a Pareto optimal way.

So, the consequentialist argument, by itself, does not suggest that the aim of profit-maximisation, or the more general aim of maximising the shareholders' interests, makes us count an organisation as a public corporation. Put differently, were the aim of profit maximisation absent, then we would not rescind the status of a public corporation. Rather, we would cease to view it as a good public corporation.

The Organisational Structure of the Public Corporation as a Solution to a Cooperation Problem

For the following argument, let us consider a characteristic feature of the organisational structure of the public corporation. In Section 6.2, I stated that one of the defining features of the public corporation is the legally required delegated management under a board structure. Many in the economic and legal literature believe that the relationship between the board and the shareholders is best described as an agent-principal relationship. This relationship

comes with *principal-agent-costs* stemming from the informational and motivational asymmetry characteristic of a bilateral relation between the shareholders (the principal) who hire the managers (the agent) organised by a board structure to work on their behalf. These costs are influenced both by the difficulty of ensuring that the agent is sufficiently motivated to act in the interests of the principal and by an informational advantage that the agent has over the principal regarding the production, which makes it harder for the principal to estimate the value of the agent's work. The principal-agent relationship and the costs associated with it are characteristic of a vertical hierarchy in which the principal both owns and controls the corporate entity. This view is often associated with the conception of a firm as a nexus of contracts.¹²

Does this argument support the ESG-rule? If the best way to model the relationship between the board and the shareholders is as a principal-agency relationship, then the associated cooperative problem that must be solved is bringing the interests of management in line with the interests of shareholders, which, in a stock market with transferable shares, comes down to maximising profits. This argument assumes that if the aim of maximizing profits is implied by the legal conditions of the ESG-rule, then the aim should itself be added to the set of antecedent conditions of the status of the public corporation. This argument may be strengthened further. Kraakman and Hansmann argue that transferable shares and management organised under a board structure are just as much part of the essence or nature of a public corporation as is the fact that the corporation is owned by the shareholders. This ownership can be legally interpreted as a fiduciary relationship which, in a liquid stock market in which investors aim at maximizing their returns, implies the duty to aim at profits. But this does not actually imply the ESG-rule for two reasons. First, in this context, the aim of profit maximisation is implied by the conditions that are in the ESG rule's antecedent. Profit maximisation is implied by them, rather than being part of the set of conditions that establish the status of the public corporation itself. The logical relation between the public corporation and the aim of profit maximisation is such that the aim is a necessary condition, entailed by being a public corporation, rather than an aim that is a sufficient condition for something counting as a public corporation. As a necessary part of what it is to be a public corporation (being implied by the conditions that set up the status of a public corporation) we can use it to categorise the institution in

¹²See Stout & Blair (1999). Eisenberg (1998) argues that the nexus of contracts conception practically eliminates the concept of the firm, since it is no longer possible to distinguish between the firm and the market, which is, in effect, also a huge system of contracts.

question. We could not do the same if we thought of the aim as being merely a sufficient condition.

Secondly, even if we were to assume that the ESG-rule is true and to assume, further, that there is a public corporation which does not aim at making profits, then we would simply say that the principal-agency relationship is defective in this very case, instead of concluding that it doesn't exist. Here, again, the logical relationship is important. People argue that managers have a fiduciary duty towards shareholder even if they don't honour it, simply because they are part of a public corporation. If by not honouring it, they would thereby rescind the status of their group agent, then this argument could not be made. Thus, normative arguments to the effect that there is a fiduciary relationship cannot be used to argue for the ESG-rule, since one can fail to satisfy a duty while still being under it. If, as a response to this argument, we would change the ESG-rule such that we would include not the fact that the corporation must aim at profit maximisation, but simply that it ought to aim at it, then the rule would not be well-formed. Duties and rights are best placed in the deontic status of the rule, not in its antecedent condition.

This means that we have found no support for the aim of profit-making among the conditions that establish the status of something being a public corporation. Hence, we have found no support for the ESG-practice. We will now look whether the shareholder governance practice is better understood if the aim of making profits or maximising shareholders' interests appears in the corporation's deontic status.

6.3.2 Deontic Shareholder Governance

Consider the following constitutive rule.

Definition 6.3.2. Deontic Shareholder Governance (DSG-rule):

If an organisation X satisfies the legal conditions highlighted in Section 6.2, then X counts as a public corporation, and, therefore, its board of directors ought to decide in the interests of its shareholders.

The DSG-rule connects the aim of profit-making closely to the obligations that managers are under by locating that aim in the deontic status of the board of directors.¹³ (Typically, the Z term that, in the DSG-rule consists only of the managers obligations towards their shareholders, will also contain other rights and duties, like being able to sue someone on behalf of the

¹³It might even be the case that the DSG-rule applies directly to the corporation and only indirectly to the managers.

corporation, borrowing money from a bank on behalf of the corporation, and having to annually lay out the finances to the shareholders). I have taken the ‘ought’ here to apply to managers and not to the corporation as a whole because this is how the business ethics debate about shareholder governance is conducted. It is typically concerned with the relationship between managers and shareholders or with incentive schemes targeting managers. Later, in Section 6.3.4, I will show how this practice, which can also be understood as providing an external account of the role of managers in a public corporation, translates back to the corporate entity as an independent locus of agency and can be translated into a duty which applies to the organisation itself. The existence of the practice thereby explains how characteristic commitments and obligations are, so to speak, baked into the organisation by accompanying and supporting the function that the entity is supposed to have as a matter of collective acceptance.

To understand the nature of the ‘ought’ in the status of the DSG-rule, one must consider its role in the collective agent’s overall mental economy. A central rational requirement is *enkrasia*. *Enkrasia*, very roughly, states that the following combination of mental states is irrational: an agent believes that it ought to *X*, while not intending to *X*. I understand the ‘ought’ which features in *Enkrasia*, following Broome (2013), to be normative, all-things-considered, forward-looking and subjective. The ‘ought’ that appears in the DSG-rule, on the other hand, is *not all-things-considered*, but it is *normative* in that it contains part of the answer that features in the all-things-considered ought of *enkrasia*. Thus, it is part of the answer that leads a rational agent to action, that explains why the agent ought to do something. The source of the normativity of the DSG-ought will be discussed below. It explains why one ought to adopt shareholder governance. It can be prudential, moral or due to the internal organisational structure of the corporation itself.

In what follows, I will consider three different strategies to explain this ‘ought’.

Corporate Survival

The first argument is prudential and from the perspective of the corporate entity as a whole. It is effectively summarised in Milton Friedman’s seminal paper ‘The Methodology of Positive Economics’.

Confidence in the maximization-of-returns hypothesis is justified by evidence of a very different character. [...] unless the behaviour of businessmen in some way or other approximated behaviour consistent with the maximization of returns, it seems

unlikely that they would remain in business for long. Let the apparent immediate determinant of business behaviour be anything at all — habitual reaction, random chance, or whatnot. Whenever this determinant happens to lead to behaviour consistent with rational and informed maximization of returns, the business will prosper and acquire resources with which to expand; whenever it does not, the business will tend to lose resources and can be kept in existence only by the addition of resources from outside. The process of “natural selection” thus helps to validate the hypothesis — or, rather, given natural selection, acceptance of the hypothesis can be based largely on the judgment that it summarizes appropriately the conditions for survival.¹⁴

Although not made explicit in his argument, the idea that a characteristic feature of public corporations is the existence of a liquid, secondary stock market, comes into play when a business, due to its profitability, is able to acquire financial resources. If we assume that participants in the stock market aim to maximise their return on investments, they will invest in those businesses which they deem to be most likely to increase shareholder value and profits. The former case allows them to sell their stocks later at a profit, while the latter case often leads to a transfer of corporate profits to shareholders by way of dividends paid out per share owned. Public corporations listed on the stock market persist and acquire resources only if they can bring stock market participants to invest in their shares. The cooperation problem they thereby face is one of trust. To get people to trust in their corporation as a vehicle of maximizing returns, they have to adopt the aim of maximising profits and act in ways coherent with it. These survival conditions, both those due to competition in the sale of goods and services and those that concern the acquisition of financial resources through the stock market by solving the trust problem do not yet explain the DSG-rule, since the DSG-rule is concerned with ‘oughts’ that managers find themselves under.

There are two ways that these survival conditions can translate to the ‘oughts’ that managers find themselves under. First, corporations often tweak the incentives of managers through schemes of participation in profits.¹⁵ But the stock market can also put pressure on the managers directly through hostile take overs. These happen when investors see the potential

¹⁴Hausman (1994, 158).

¹⁵See, for example Proctor and Gamble, *Structure and Governance - Website of Proctor and Gamble* (n.d.), who state in their "corporate governance overview" that "Our Executive Share Ownership Program requires senior executives to own shares of company stock and/or restricted stock units valued at eight times base salary for the CEO, and five times base salary for the other senior executives. Non-employee directors must own

to increase profits with a firm's presently available resources by changing its strategy. This can give them an incentive to take over the company by removing the management team and substituting it with one that is conducive to greater profits. As Heath argues, such hostile takeovers would succeed in most market societies since strategies to protect managers from these raids or from being fired by shareholders are typically prohibited.¹⁶

Shareholder Governance and Fiduciary Duties

Historically, shareholders have been viewed as the owners of corporations. On that view, shareholders employ directors to manage their assets in their interests. This was noted by Berle and Means in the early 20th century.

Wherever one man or a group of men entrusted another man or group with the management of property, the second group became fiduciaries. As such they were obliged to act conscionably, which meant in fidelity to the interests of the persons whose wealth they had undertaken to handle.¹⁷

It is this idea that a principal, who owns assets or equity, comes to a voluntary agreement with an agent who will manage her assets in her interest, that is supposed to explain why the managers in a corporation generally, and the board of directors in the specific case of the public corporation, have fiduciary duties to their shareholders. In the famous paper 'The social responsibility of managers is to make profits' Friedman builds on this argument, claiming that '[t]he whole justification for permitting the corporate executive to be selected by the stockholders is that the executive is an agent serving the interests of his principal.' Appeals to the principal-agency relation try to morally and legally ground both the *existence* and *uniqueness* of the fiduciary duties which managers owe to shareholders. In law, fiduciary relationships come with a set of duties that the fiduciary owes to her principal. The duty of care requires that directors, when making a business decision, inform themselves to a sufficient degree, assessing the information with a 'critical eye' so as to protect the interests of the principal. The duty of loyalty requires that there are no conflicts of interest that prompt the fiduciary to abuse their position to further their private interest.¹⁸

company stock and/or restricted stock units worth six times their annual cash retainer. These compensation programs help to ensure the alignment of the interests of our senior executives and directors with shareholders."

¹⁶Heath & Norman (2004)

¹⁷Berle & Means (1932, 336)

¹⁸Further duties include the duty of prudence, confidentiality and disclosures.

In what follows, I will present three arguments that have been made for both the existence and uniqueness of the fiduciary duty between shareholders and the board of directors.

Fiduciary Duties and Ownership

Some people have argued that legally, shareholders are the only stakeholder group towards which executive directors have fiduciary duties and the only stakeholder group which has the right, within the organisation, to sue the board of directors for the violation of those duties. Furthermore, shareholders alone are allowed to elect the members of the board of directors and let them decide on fundamental corporate decisions like mergers and acquisitions. These rights have given some the impression that shareholders are *legally* the owners of the corporation, since on the surface they already seem to be enjoying some of the characteristic ownership rights over it.

Kraakman & Hansmann (2017), for example, take ownership of the corporation by investors of capital to be the defining feature of the public corporation.¹⁹

Despite this, closer inspection reveals that these rights are all limited, most strongly by the business judgement rule, which protects every decision by the board as long as that decision can be rationalised as serving the corporation as a whole, rather than the specific interests of shareholders.²⁰

¹⁹See Ludwig (2017a), who follows Kraakman and Hansman in their assessment. Eisenberg (1998), on the other hand, argues that even if shareholders have characteristic ownership rights, this does not imply that corporation should be managed in the owners' interests.

²⁰In the United States, those who argue that the board of directors should primarily serve the shareholders of the corporation typically allude to two legal aspects of corporate law: (1) derivative suits for breach of fiduciary duties and (2) shareholder voting rights.

Regarding (1): Derivative suits cover cases in which the board of directors does not act in the interests of the corporation, for example, cases of self-dealing. Since the board, under normal circumstances, should represent the interests of the corporation itself, when the board members fail to live up to the fiduciary duty that they owe to the firm, the shareholders are allowed to step in and sue them on behalf of the corporation. Importantly, here, the corporation is treated as a legal person with independent interests which are violated exactly when board members engages in self-dealing and the abuse of corporate opportunities, 'reaping profits from personal business ventures that either are in the same line of business as the firm's, or became available to them because of their corporate position.' See Stout & Blair (1999, 299). Importantly, although derivative suits can only be brought forward by shareholders and directed at the members of the board of directors, they have to be filed on behalf of the corporation as an independent legal entity. The fact that only shareholders can sue on behalf of the corporation might suggest that derivative suits protect the shareholders' interests. As Stout and Blair argue, this is not so, since shareholders can't sue a corporation for the violation of *their* interests – for example,

Even if we would accept that a corporation can be owned by its shareholders, we might still ask why ownership rights over equity transfer onto a legal body that involves the interests and actions of other people, like a corporation. As Stout & Blair (1999, 261) convincingly point out, it seems rather odd that someone can have ownership rights over a corporation's assets once the intellectual capital and experiences of its workforce are included. Moreover, treating people's actions as things that one can own seems to be beyond the domain of what one can morally and rightfully own. Thus, it is no surprise that the idea that shareholders own the public corporation is also in deep conflict with the legal reality in which the corporate assets are owned by the corporation itself and, therefore, can only be used and sold by the board of directors. Legally, the shareholders own the stock, rather than owning the company.²¹

Fiduciary Duties and Shareholder Vulnerability

Another argument which turns not on the notion that shareholders own the corporation, but on the notion that the equity that shareholders forward to the corporation ought to be protected from mismanagement and exploitation, tries to establish that the relationship between shareholders and managers is special for being particularly vulnerable. It starts from the premise that if shareholders and managers were to try to come up with a contract fixing the interests and demands, no such contract would be practically available. The decisions which managers face are sufficiently complex and unpredictable because more often than not they happen on a trial-and-error basis so that

their interest to maximise corporate profits – but rather for the violation of the corporate interests more generally. This is supported by case law, which suggests that the board of directors are trustees of the firm itself. Stout & Blair (1999, 303) write that ‘judges have [not] sanctioned directors’ decisions to use corporate funds for charitable purposes; to reject business strategies that would increase profits at the expense of the local community; to avoid risky undertakings that would benefit shareholders at creditors’ expense; and to fend off a hostile takeover bid at a premium price in order to protect the interests of employees or the community. As these examples illustrate, modern corporate law does not adhere to the norm of shareholder primacy. To the contrary, case law interpreting the business judgment rule often explicitly authorizes directors to sacrifice shareholders’ interests to protect other constituencies.’

Regarding 2: Shareholders’ voting rights are similarly narrow. They can vote only on a few issues such as mergers and the makeup of the board. But not only are their interests, due to the sheer number of shareholders in publicly-traded corporations, diverse to the point where no dominant interests or directions emerge. They also typically have only veto rights and often can only suggest new board members at the directions of the existing board.

²¹Stout (2001, 1191)

it is not feasible to specify in advance how to respond to a wide range of future contingencies.²² With no contract available and the whole investment of the shareholders potentially ‘placed at hazard,’²³ the fiduciary duties of managers towards the shareholders take the role of *gap fillers* which could not otherwise have been taken by other forms of contracts. This argument tries to establish that shareholder governance solves a problem that arises in long term, contractual relationships, in which success conditions are difficult to define, while at the same time highlighting that shareholder governance also protects the property interests of the investors in a public corporation. It rests on the assumption that the relationship between the corporation and its workers is less vulnerable and easier to protect by courts, which is ultimately an empirical question.

Fiduciary Duties and Promises

A third line of argument hones in on an implicit voluntary agreement between shareholders who ‘advance their money to business managers on the condition that it be used in accordance with their wishes.’²⁴ This condition is supposed to give rise to a kind of contract or promise. Ownership still plays a role, but only in facilitating the contract. According to these authors, it is this promise or contract to use the money in the interests of the shareholders which makes managers the agents of the shareholders.²⁵ This point is concerned with the intrinsic moral value of promises and agreement.

Typically, for there to be an agreement of the form cited above, the parties to that agreement must, in paradigmatic cases, both openly express their willingness to be part of it and must be able to negotiate what that agreement consists of.²⁶ Clearly, this is not the case in a public corporation with transferable shares which are often purchased from previous owners who might have held them for only a very brief time. Even when these shares are issued for the first time by the company, the ‘shareholders expect to be treated as ‘investors’ [...] and expect corporate managers to consider a wide constituency when making corporate decisions’²⁷ or simply have different,

²²Macey (1991, 27)

²³See Boatright (1994, 395) quoting the influential work of Williamson (2007).

²⁴Hasnas (1998, 9-10)

²⁵See Goodpaster (1991), who discusses managers promising shareholders a return on their investment or Dodd Jr (1931), who has, through an extended debate with Berle (1932), also framed the relationship between shareholders and managers as one of trust.

²⁶See Gilbert (1992) who bases her idea of joint actions on agreements and Boatright (1994) who suggests that the normal conditions which are required for a relationship to count as a contract are not met between the shareholders and management.

²⁷See (Boatright 1994, 397-98), who quotes from a quantitative study done by

incompatible interests and expectations towards the company.²⁸ Similarly, shareholders are often represented by an intermediate entity themselves, as when a pension fund invests in a company on behalf of its members.²⁹ So, it is not clear how one would arrive at the content of the agreement that Hasnas and others assume to be at the heart of the shareholder management relation. Nor is it clear whether any special agreement exists which justifies speaking of a fiduciary duty that management owes only to shareholders and not several fiduciary duties owed to different stakeholder groups.³⁰

6.3.3 Anchoring the DSG-rule

We have now explored some of the reasons offered in the business ethics literature for why managers should aim at profit maximisation. These reasons were prudential reasons related to the continued survival of the company, but also moral in nature. Among them are not just those that try to characterise the relationship between managers and shareholders as one between promisor and promisee, but also those that argue why shareholder governance is the most efficient solution to characteristic agency or cooperation problems and maximises the overall social welfare (Section 6.3.1). We also looked at whether these arguments give rise to the existence of a unique fiduciary relationship between managers and shareholders.

Before the backdrop of the DSG-rule, we can understand these arguments as arguments why one should participate in the collective acceptance of the DSG-rule. Put differently, these are arguments that explain what realises or puts the DSG-rule into place rather than, say, the ESG-rule. They are, Brian Epstein has argued, question about what anchors the practice.³¹ For these reasons to be an answer to the anchoring inquiry, people must participate in the rule for those reasons and enough people must do so. Since we understand shared acceptance via Bratman's account of shared intentions, we have more to say about the role that these arguments in favour of shareholder governance play. First, they are important for each participant in the shared acceptance of the DSG-rule to *persistently* play their part. Thus, people must assume that the conditions which validate the different normative arguments in favour of the practice persist. Since Bratman is individualistic

Soderquist & Vecchio (1978, 840).

²⁸Eisenberg (1998, 833).

²⁹See Eisenberg (1998, 824).

³⁰Arguments that all participants to the public corporation would come to an agreements under the veil of ignorance to favour shareholder primacy seem to be at odds with the idea that the interests of employees will often times be in conflict with the goals of shareholders.

³¹See Epstein (2015, 2019).

about the reasons why someone might participate in a shared intention, we may also note that different people who accept the DSG-rule and partake in the practice might do so for different reasons. Some might believe that profit maximisation is morally required, while some might do so for fear of losing their jobs or for reasons connected to higher pay in case where the entire corporation is doing better and their pay is linked to the overall corporate performance. Others, identifying with the corporation and being invested in it, might care deeply about its survival and for this reason participate in the collective acceptance of the DSG-rule.

Another important feature of the DSG-rule is that it gives rise to a practice that does not stop with the managers. Part of its normative force is derived from being shared not just by managers of corporations, but among wide parts of society. Indicative of this is that the above arguments in favour of shareholder governance have had an enormous influence in the academic literature. One of the most widely read introductory books on economics by Mankiw (2014), for example, posits profit maximisation as a corporation's defining aim. As Jensen (2001, 8) says, 'Most economists would answer simply that managers must have a criterion for evaluating performance and deciding between alternative courses of action, and that the criterion should be maximization of the long-term market value of the firm. This value maximization proposition has its roots in 200 years of research in economics and finance.' In law, Kraakman & Hansmann (2017) have famously argued for a convergence of international law towards shareholder governance. Management theorists Jensen & Meckling (1976), Jensen (2001) and Drucker (1984) have also argued for shareholder governance. These papers have been quoted collectively over two hundred thousand times, which hints at their enormous impact, even though, as I have tried to show, each specific argument might not be convincing in itself.

But the widespread uptake of shareholder governance has not been confined to academia. Famous case law similarly supports the idea that corporate CEOs should aim at profits. Take, for example, the notorious 1919 case of *Ford v. Dodge*,³² where the Chief Justice Ostrander justified his ruling in the following way.

A business corporation is organized and carried on primarily for the profit of the stockholders. The powers of the directors are to be employed for that end. The discretion of directors is to be exercised in the choice of means to attain that end, and does not extend to a change in the end itself, to the reduction of profits,

³²Ostrander (Mich. 1919)

or to the non-distribution of profits among shareholders in order to devote them to other purposes.

This quote captures some of the core idea underlying the DSG-practice, justifying it on the basis of the relationship between managers and shareholders. It clearly shows how the practice is not only grounded in the interconnected beliefs of the managers of public corporations and academics, but also in those beliefs that comprise the body of law. Although this broad acceptance might have recently changed,³³ we can still assume that it is upheld widely today, given the current canon that is taught in introductory economics courses.

Now, let us see how the DSG-practice might impact the agency of the public corporation.

6.3.4 The DSG-practice as a Practical Identity.

Let us return to a question which I have touched upon before, namely, how does the practical identity – the DSG-practice specifically – shape the corporation’s mind and the point of view? Especially, how does it influence that point of view, if the managers themselves don’t participate in it. To see this, let us look at a concrete example. Take the managers at a public corporation such as Daimler, which sells cars, motorbikes and buses. Assume that the managers, and the board in particular, collectively accept the DSG-rule because of some of the reasons above: Maybe they participate in the corporation’s profits via shares they own, maybe they believe that this way the corporation is more efficient or maybe they think that there is an implicit promise underlying the relationship to their shareholders.

Having collectively accepted the ought-belief ‘aiming at maximizing Daimler’s shareholders’ interests’, when deliberating from the perspective of the corporation either by themselves or collectively, the managers will give the most weight to the interests of the shareholders. They will intend the means to the shareholders’ ends, and they will try not to form plans and goals which are inconsistent with these interests, allowing the different sub-plans of the corporation which are debated and enacted, to mesh. Of course, this also means that if someone brings up values that are deeply in conflict with the aim of making profits, like producing all Daimler cars in Germany or stopping the sale of extremely energy-inefficient SUVs, then these considerations will be blocked in deliberation and will be filtered out. Practically speaking, this means that the aim of maximising the shareholders’ interests will shape their deliberation. It will be a constant element in dwelling on answers to the question of what Daimler ought to do, which means that it will be part

³³See Stout (2008) on why Dodge v. Ford has remained a surprising influence.

of the perspective of Daimler that the managers adopt when they aim at self-understanding on behalf of the corporation. Importantly, for it to be part of that perspective, they do *not* have to vote or agree on it, as List and Pettit would have it. It is sufficient for the 'ought' to play its functional role as a reason in the explanations of what Daimler ought to do and, thereby, to become part of the perspective that is assumed by the managers when they aim at making sense from the company's perspective. But, of course, including the normative consideration in their deliberation effectively will require the presence of a shared policy of giving weight which would provide a collectively agreed-upon interpretation of what Daimler ought to do.

Now, let us assume that, through a decision mechanism that demands that they vote on the premisses of a set of arguments that include the problematic environmental effects of pick-ups and other extravagant cars, the board of directors conclude that they should stop making cars which drive fast, which are energy inefficient but attractive to their customer base. Having formed this intention, the managers will now communicate their decision to cease making and selling big, energy-inefficient cars to the shareholders and the media, a sufficient number of whom collectively accept the DSG-rule and participate in the resulting practice. Their reaction will be, given their participation in the DSG-practice, highly adverse. Shareholders and the media, on grounds of participating in the DSG-rule, will not accept the reasons brought forward by the board, since being part of a practice that fixes the 'oughts' that managers are under in the way the DSG-rule suggests, will block environmentally-friendly considerations from having any effect if they undermine profits or directly threaten the survival of the company. Thus, the DSG-rule will not only influence the corporation's internal deliberation: if accepted in a broad enough fashion, it will also put pressure on what decisions are reasonable as understood by agents external to the corporation and put pressure on the members this way. The point here is that the members of the corporation need not even participate in the collective acceptance of the DSG-rule for people to attribute the 'ought' of maximizing shareholders' interests to the perspective of the corporation by way of representing the public corporation as being under the 'ought' fixed by the DSG-rule. Anticipating this, will then influence what managers themselves will find intelligible from the perspective of their public corporation. In this way, a practice which, like the DSG-rule, is anchored by way of collective acceptance spanning across wide parts of society, influences the 'mind' and actions of public corporations. It influences both what members of the corporation find intelligible for the corporation to do, but also what agents external to the corporation find intelligible for the corporation to do.

6.3.5 Objections against the External Practical Identity

There are two final objections against the overall strategy that I have not yet addressed. One concerns the idea that shareholder governance must be understood as a social practice rather than a set of individual beliefs. The other points at the demandingness of parts of my account of shared intentionality, specifically, the demandingness of Bratman's account of shared intentions. I will now turn to each of these.

The Demandingness Objection

A critic might not be happy with my characterisation of the social practice understood through constitutive rules, in case the shared intentionality is fleshed out by Bratman's concept of shared intentionality. Bratman's account of shared deliberation requires common knowledge among the participants of the relevant conditions that constitute the functional profile of the shared intentional activity. It is unlikely that every or most participants in the shared activity meets this condition, since it might not be clear who participates in the social practice or whether that person knows that the others know that they participate, and so on. Furthermore, there will be many people who are also relevant to the public corporation, like some of the stakeholder groups, political parties and members of the media who participate in an incompatible practice, or in no practice at all, relevant to the function of the firm to begin with.

The first way to respond to this objection is to simply accept it. The critic is absolutely right that, in reality, the social landscape that shapes the firm's function is much more diverse and heterogeneous than we have assumed. Nevertheless, I want to suggest a way that we can hold onto the claim which I have made, albeit with somewhat weakened premises. Consider how Bacharach et al. (2006) and Gauthier (1986) treat the case of cooperation among people who are not unanimously inclined or disposed to cooperate. In these situations, cooperation might still happen if enough people are willing to cooperate and this is out in the open and if the cooperation's expected pay-off is greater than the expected losses in case participants decide not to cooperate. Perceived participation in a practice might work in a similar way. If people think that the reasons to join a practice are particularly strong in a certain situation, then they might assume that a substantial number of people who find themselves in similar circumstances also end up endorsing the practice and playing their role in it, even though they don't know exactly who specifically participates in it.

Objection from Individually Held Beliefs

Why, one might ask, should we think of shareholder governance as a practice, rather than a set of beliefs that everyone either believes or doesn't believe *individually*?

To see why we should, we must go back and consider the role played by constitutive rules and the practice which they give rise to. Constitutive rules, such as those that allow dollar bills to count as money or those that make the red light at an intersection count as a traffic light with the relevant deontic powers associated, help to solve coordination and cooperation problems.³⁴ Money solves the problem of the coincidence of preferences which occurs in a trade economy, while traffic lights solve the coordination problem of how to safely pass through an intersection. As we have seen, shareholder governance is supposed to solve a number of coordination problems, among them the different variants of the principal agency problem between shareholders and managers, and the general allocation and efficiency problems in a market environment. Economists, social scientists and philosophers as early as Lewis (2008) have shown these problems to be insurmountable if we reason individually and strategically on the basis of individual expectations and desires.³⁵ Even something as simple as crossing an intersection safely is impossible if there are more than two ways to do it, which there are – I could go first while you wait or you could go first while I wait. The problem arises if we both go at the same time or don't go at all. It is easily solved if we introduce a practice according to which a traffic light governs when each driver is to go. The same argument holds for the practice of shareholder governance. If it is to have the benefits associated with it at all, it must be collectively accepted. It cannot be just a set of beliefs held by each individual independently of everyone else.

The second argument is similar in spirit to those advanced by Margaret Gilbert, the philosopher who first introduced the notion of shared intentionality in her book *On Social Facts*:³⁶ There, she argued that a simple phenomenon such as walking together cannot be made sense of if we think of two people merely holding individual beliefs about walking to the same destination with the other person. These individual intentions can be satisfied if the 'partner' is kidnapped or forced along by a gun, or would be satisfied if, at any point, one of the walkers speeds up, leaving the other behind.³⁷ If, on the other hand, two people agree to walk *together*, they are responsive to

³⁴See Hindriks & Guala (2019), for example.

³⁵See also Bacharach et al. (2006) and Gauthier (1986).

³⁶See Gilbert (1990, 1992, 2006).

³⁷See Bratman (1992, 1993) for the kidnapping example.

each other, they should tell the other whether they want to exit the shared activity, and can make demands on the other person such as demanding that the other walk a little slower.³⁸ That the two people have a certain shared intention also means that incompatible ends are filtered out. This functional profile, connected to shared intentions, cannot be explained by individual intentions that are not socially glued together.

If the kind of deliberation governed by the DSG-rule was not collectively accepted, then incompatible aims would not be filtered out and the aim of putting shareholders' interests first would not be stable, being liable to be dropped at any time. Members of the corporation and outsiders could not engage in collective deliberation that takes the function of the public corporation as fixed, and the cooperative problem of trust between shareholders and managers could not be solved. For these functions, it is required that the DSG-rule is a practice, rather than a belief held by each participant independently. Thus, it is implicitly assumed by those arguing for shareholder governance to be the defining practice of the role of managers of the public corporation. Finally, aims like profit maximisation or, say, balancing the interests of all stakeholders, could not constitute the external practical identity and, thus, could not play the diachronic rational function played by a stable identity.

6.3.6 Further Remarks on Social Practices

It is an empirical question how the 'ought' in the DSG-rule is interpreted and how widely it is shared. Recently, an influential set of institutions managing a significant part of the investments of the capital in circulation has demanded that corporations become carbon neutral. In other markets, such as the market for drilling oil, the investors' interests might be more traditional, concerning the growth of corporate value or profits. Although even here, shareholder activism focused on advancing policies which try to address climate change, even in the fossil-fuel production market,³⁹ can weaken the DSG-rule or change how it is interpreted. For example, they may affect whether and which moral or value constraints on profit maximisation are adopted.

Markets also differ in how competitive they are, which, in turn, also in-

³⁸Admittedly, this is a combination of both Bratman and Gilbert's theory that Alonso (2009) and Rachar (2021) have recently argued for.

³⁹*Exxon loses board seats to activist hedge fund in landmark climate vote* (n.d.) reported that Exxon, one of the staunchest fighters against attempts to curb greenhouse gas emissions, had to accept three board members suggested by an activist hedge fund which aims to introduce stronger measures cutting carbon emissions.

fluences how the ‘ought’ will be interpreted, whether it is easily outweighed, for example, in those markets that are not competitive or whether it is taken to outweigh other considerations, as in competitive markets. In highly competitive markets, the argument from survival is stronger than in markets where firms hold a monopoly. Later, we will consider another practice that is known under the heading of stakeholder governance. Here, again, the same thing which is true with regard to shareholder governance also holds true for stakeholder governance. There might be some markets and legal environments in which stakeholder governance is the central social fact that fixes a corporation’s external practical identity.

These brief thoughts are meant to illustrate the complexity involved in characterizing the practical identity of institutions, even if we focus solely on the public corporation and its function in our society. There is no one practice by which we may characterise the external practical identity of all public corporations. Nevertheless, it is likely that there are markets which are characterised by the DSG-rule, for example, because they are highly competitive or because the CEOs and shareholders are, due to their economic background, convinced by the moral arguments for shareholder governance. When the DSG-rule shapes the mind of the corporation in such markets and is interpreted to contain a maximizing ought, then, I will argue, we are faced with a one-dimensional valuer. The following section is intended as a proof of the existence of the ODV, or at least of evidence of its possible existence. The argumentative strategy will be somewhat similar to the Sherlock Holmes method which we encountered in Section 1.7, whereby we try to amass evidence for the existence of a certain agential structure or a characteristic end in systems inhabiting a specific niche.

In what follows I present an interpretation of the DSG-rule advocated for by business ethicists which mirrors the reasoning and moral deliberation characteristic of the ODV. Furthermore, I will show how such reasoning can also be witnessed by corporations themselves and is shown with respect to the media.

6.4 Instrumental Ethics

When the DSG-rule is interpreted not as a qualified ‘ought’ which partially answers the question of what the corporation should do, but outweighs other considerations, we end up with the following understanding of the corporation, summed up by Goodpaster.

The foundations of ethics in management [...] [lie] in understanding that the conscience of the corporation is a logical and moral

extension of the consciences of its principals. It is not an expansion of the list of principals, but a gloss on the principal-agent relationship itself.⁴⁰

If the principal's aim is to make money, Goodpaster's statement translates into Milton Friedman's famous, ostentatiously-titled essay *The social responsibility of business is to increase its profits*.

As we have noted before, someone might also participate in the collective acceptance of the DSG if they take a public corporation to inhabit a highly competitive market. These perfect markets, Baumol and Black observed, 'impose vice rather than virtue',⁴¹ echoing Hume's sentiment that 'if it should be a virtuous man's fate to fall into the society of ruffians [, then] his particular regard to justice being no longer of use to his own safety or that of others, he must consult the dictates of self-preservation alone'.⁴² Competitive markets and the associated necessity to maximise profits as a means to self-preservation are one way in which the 'ought' in the DSG-rule is justified as outweighing other normative considerations. The business ethics program that was built around the DSG-rule, interpreted in this maximising way, is known under different names: *instrumental ethics*, *strategic stakeholder maximisation* or *enlightened value maximisation*. 'Its advocates employ the language of ethics to serve the goal of firm value maximization. Instrumentally ethical managers might do what is morally proper, but they do so to increase shareholder wealth.'⁴³

Jensen, one of the most influential academic voices in the business and management literature does not consider instrumental ethics problematic, but, indeed, approves of it under the name of 'enlightened value maximisation'. The kind of ethical reasoning for which he advocates is exactly similar to the one which I have explored in an earlier part of the dissertation, and which is characteristic of the OD. He justifies the underlying DSG-practice by appeal to the consequentialist argument which we looked at closely in Section 6.3.1.

Without the clarity of mission provided by a single-valued objective function, companies embracing stakeholder theory will ex-

⁴⁰Goodpaster (1991, 68)

⁴¹See Baumol & Blackman (1991, 3), quoted by Quinn & Jones (1995, 26). In their book, Baumol and Black 'contend that it is very difficult for competitive firms to behave in socially desirable ways – being kind to the environment, contributing to worthy social programs, handling redundancy humanely.'

⁴²See Hume et al. (2000, 39) and also Gauthier (1986) for the same point that moral restraint is not rational in social dilemma like situations in which no one intends to cooperate.

⁴³(Quinn & Jones 1995, 23)

perience managerial confusion, conflict, inefficiency, and perhaps even competitive failure. [...] In order to maximize value, corporate managers must not only satisfy, but enlist the support of all corporate stakeholders. [...] Enlightened value maximization uses much of the structure of stakeholder theory but accepts maximization of the long run value of the firm as the criterion for making the requisite trade-offs among its stakeholders.⁴⁴

Here, Jensen is saying that stakeholder concerns are only relevant insofar as they impact the primary objective of the public corporation, value maximisation. It is important to note that this kind of value is not moral value, but the ‘sum of the market values of the equity, debt and any other contingent claims outstanding on the firm.’⁴⁵ The reason why he thinks that a corporation that serves many masters or has many conflicts of interests will count as non-purposeful or confused is laid out in the following lines.

It is logically impossible to maximize in more than one dimension at the same time unless the dimensions are what are known as ‘monotonic transformations’ of one another. Thus, telling a manager to maximize current profits, market share, future growth in profits, and anything else one pleases will leave that manager with no way to make a reasoned decision. In effect, it leaves the manager with no objective. The result will be confusion and a lack of purpose that will handicap the firm in its competition for survival.⁴⁶

For this argument to work, Jensen must be assuming that any corporation already adds some value to society. After all, there seems to be nothing wrong with a corporation that sells a product with huge externalities – like opioids for people without chronic pain, tobacco or coal – and that does not survive and goes out of business. Survival, as I discuss in Section 6.3.2, can at best provide the corporation with prudential reasons in support of the DSG-rule. As Pearce (2019) argue, markets such as coal and tobacco have a negative net value. For these markets, there is no set of moral constraints which is in any way compatible with the objective function within the DSG-practice. This means that these agents, in deliberating and deciding in the way that Jensen proposes will cause more harm than good.

Second, even for markets with a positive net value, such as the market for cars, the kind of deliberation that settles conflicts between different interests

⁴⁴Jensen (2001, 9)

⁴⁵Jensen (2001, 12)

⁴⁶Jensen (2001, 10–11)

in the way Jensen argues for – which also mirrors the deliberation of the ODV, which results from the DSG-practice – will lead to a reduced, if not faulty, moral responsiveness unlike that of a full-fledged moral agent. This is especially pronounced in the following case.

Although Ford had access to a new design which would decrease the possibility of the Ford Pinto from exploding, the company chose not to implement the design, which would have cost \$11 per car, even though it had done an analysis showing that the new design would result in 180 less deaths. The company defended itself on the grounds that it used the accepted risk/benefit analysis to determine if the monetary costs of making the change were greater than the societal benefit. Based on the numbers Ford used, the cost would have been \$137 million versus the \$49.5 million price tag put on the deaths, injuries, and car damages, and thus Ford felt justified not implementing the design change.⁴⁷

Ford's deliberation, shaped by its constitutive aim of making money, is a case in point of how a practice that enforces a constitutive non-moral maximising end on an agent can lead to deliberation that is insensitive to morally salient features. Indeed, the deliberation is not moral at all, having an almost psychopathic ring to it.

Or, consider an article in *The Economist* on how Shell, 'the only firm to link its executive's pay to progress in reducing emissions across its operations', plans to tackle climate change.

Despite the urgency to tackle climate change, Mr van Beurden has no intention of going all in on a post-carbon future, and warns against Shell sticking its neck out too far. To explain why, he sets out a few hard truths. The first is about business itself. Shell may justifiably fear being on the wrong side of history when it comes to climate change. But it needs shareholders' support to move in the right direction. Though some investors put global warming as their highest priority, most still relish the juicy 10%-plus returns that Shell generates on capital employed in big, risky projects such as oil wells and refineries. They are wary of cleaner-energy ventures such as electricity, where Shell has taken its first steps; returns are steadier, but puny (say 4%). New energy businesses such as hydrogen and biofuels are seen as financial black holes. So Shell has to coax investors along with a mix of hard chash and

⁴⁷Leggett (1999).

prudent investments. [...] (Shell) has thought to reassure critics by earmarking most of its \$ 30bn annual capital-expenditure budget over the five-year period for fossil-fuel related projects. As for the prudence, it will only ramp up spending on its nascent power business if it can show that returns come close to those of oil and gas. Investors wanting more ambitious climate strategies can put their cash into clean-tech companies instead.⁴⁸

This case again exemplifies the same kind of deliberative logic that Jensen has argued for, but this time from an influential economic magazine. Although environmental interests are taken into account, they are only relevant insofar as they are held by the shareholders or do not negatively effect their overall profits. I could go on and on with meat producing corporations venturing into vegetarian products but otherwise factory farming their meat, to corporations adopting a range of ecologically farmed products while also continuing to produce environmentally harmful products. The point I want to make is clear. In environments like competitive markets or with shareholders seeking mostly profits, public corporations will be characterised by something like the DSG-rule, which, in turn, shapes their deliberation, as seen in the case of Shell and Ford. This kind of deliberation is non-accidental, but indeed accepted by media outlets which, like *The Economist*, normally champion those solving climate change problems, but accept excuses that are grounded in the DSG-practice.

As these cases show, the ends which characterise the external practical identity of corporations do influence their capacity to reason morally and to respond to moral reasons by changing what decisions and rationales for these decisions are intelligible from their own point of view as well as to other agents who interact with them. This is already an important observation, which shows that the ends of corporate governance fixed by a practice such as the DSG-rule influence the moral capacities of corporations, thereby also influencing the kind of moral standing which we should give to corporations.⁴⁹

The Public Corporation and Self-Intelligibility

There are two ways in which group self-intelligibility might be influenced by the DSG-rule. The DSG-rule, at the level of the group's practical identity, puts pressure on the corporation to do what is rational or intelligible in light of the profit-maximising aim that the rule contains. But achieving self-

⁴⁸Schumpter (July 6th 2019)

⁴⁹Contrary to what Moriarty (2017) claims, who believes that discussions about corporate governance do not have any actual consequences for business ethics.

intelligibility itself might be influenced by the constitutive, unalterable end of profit-maximisation.

Suppose that the members of the public, oil-producing corporation *Shill* participate in the DSG-rule, perhaps because of the prudential arguments explored in Section 6.3.2 or because of Jensen's value-maximising view. In this case, the way to interpret and implement the 'ought' in the DSG-rule does not leave much room for the kind of open future that is relevant to group self-intelligibility, as I argued in Section 5.6.2. *Shill* lacks the kind of practical deliberation required for moral comprehension, and will not be able to choose in a way that is responsive to the moral facts as they present themselves from their individual perspective. Thus, the group will not meet Wallace's conditions for moral responsibility and it will be incapable of a fully autonomous will as specified by the constitutive account of group agency. If it wants to act morally, it must transcend itself, its own peculiar commitments, projects and its external practical identity and the DSG-practice that is constitutive of it. As Velleman puts it 'self-transcendence in these cases always involves some irrationality'.⁵⁰

Imagine, then, another public corporation listed on the stock exchange, *Shall*, which invests in renewable energy, because the founders are convinced that the market for renewable energies is not too competitive and does not create high external costs. Thus, they participate in the DSG-practice, because of the welfare argument.

This allows us to build a bridge back to the intuition with which we began this dissertation. There, we observed that corporations are two-faced, in that they both seem to be, in principle, capable of moral agency and, at the same time, are set up so as to pursue a pre-specified goal. The tool metaphor is meant to convey that an agent cannot choose its own aims, which is one of the required features of moral agency. The corporation *Shill* is more like a tool than an agent. *Shall*, although having a maximising non-moral aim, has, by comparison, more leverage on interpreting the social practice that it is part of. Even though its practical identity is also shaped by the DSG-practice, the justification for the 'ought' in the rule is different. This not only means that the reasons for which the members of *Shall* participate in the DSG-practice are different. It also affects the guiding force of the DSG-practice and the practical deliberation that it makes possible. Compared to the DSG-practice based on prudential reasons, it allows for different exceptions and different qualifications to the end of making profits in the face of conflicting moral obligations. If the DSG-practice within a social structure is based on the consequence or welfare argument from Section 6.3.1, then its participants can

⁵⁰See Velleman (2006, 13).

adopt Heath's market failure approach to business ethics, which means that a corporation will counteract the external costs it causes through pollution, for example, by investing into more expensive technology to counteract or prevent pollution.

The DSG-practice and Amoral Agency

Are these cases illustrative of agents which are amoral or merely immoral? The answer depends on our concept of responsibility. On the Strawsonian understanding of responsibility which I have adopted, corporations shaped by the DSG-practice will be unable to be a proper participants in moral relationships. They will not be able to question certain ends and goals that define them, as they have no control over them and no way to adjust them if necessary, since the interpretation of the DSG-practice might be beyond their reach. In Pettit's own words, they cannot 'offer an interpretation of their own purposes and goals',⁵¹ since some of these are not, strictly speaking, their own. Under this view, these corporations are not full moral agents. They are, to a certain extent, amoral, although how much exactly is a question that can only be answered by looking at how exactly the DSG-practice is spelled out. Shall, whose DSG-practice is anchored by means of the survival argument, seems more like an amoral agent that we can punish and blame, but with whom we do not want to participate in moral relationships. Shall, on the other hand, seems somewhat capable of participating in these relationships, at least temporarily.⁵²

Given that economic conditions change, we might generally not want to include even corporate agents like Shall in our practice of responsibility, since we cannot monitor their social circumstances closely enough to know when to meet them on even terms, as one moral agent among others, and when not to.

6.4.1 Stakeholder governance

Support for understanding the function of the public corporation as one that should maximize profits remains broad. But there is at least one competing view of the function of the corporation, already articulated in the early debate between Berle and Means in the beginning of the 20th century and

⁵¹(Pettit 2017, 19)

⁵²See de Haan (2022) for a view of corporate agency that argues that corporate agents are always moral agents and that something like the DSG-practice can at best temporarily excuse them from the practice of responsibility.

which gained popularity in the 1980s with the emergence of stakeholder governance.⁵³ This view has it that the public corporation ought to balance its stakeholders' interests as the interests of just one interest group among many. We can express this view as a constitutive rule in the following way.

Definition 6.4.1. *Stakeholder Governance (StG-rule)*

If an organisation X satisfies the legal conditions highlighted in Section 6.2, then X counts as a public corporation, and, as a public corporation, its board of directors ought to balance its stakeholders' interests.

This rule by itself does not dictate who the stakeholders are and how their interests should be balanced. Typically, stakeholders comprise the shareholders, employees, the community, suppliers, and customers, and sometimes also creditors, the government, and competitors. One way to balance the interests of all is to give most weight to the interests of the shareholders, which would transform the StG-rule into the DSG-rule. Over the years, some convincing arguments in favour of this conception of corporate governance have been put forward. For illustrative purposes, I will discuss an economical argument. The argument is similar in spirit to the one that justifies shareholder governance on the grounds that it solves the cooperation problem of the principal-agency relationship between shareholders and managers.

The argument for stakeholder governance begins with the observation that if the firm-specific investments of the different stakeholder groups (for example, employees and creditors) are each essential to the production of the good that is to be sold by the firm, the cooperative situation is not most accurately depicted as a principal-agent problem, but as a *team production problem*. Alchian and Demsetz define team production as 'production in which 1) several types of resources are used [...] 2) the product is not a sum of separable outputs of each cooperating resource [and] 3) not all resources used in team production belong to one person.'⁵⁴ If a firm's production of a good or service meets these three criteria, it is impossible to clearly separate each members' or stakeholders' contribution to the outcome. Each individual's input is required and essential to the final product. This makes it difficult – if not impossible – to determine how valuable each members' work was. If, in such a situation, the allocation of profits from the team activity is fixed at the start, each member has an individual incentive to unilaterally work less

⁵³One could also go further back in time, to the first corporations on American soil, whose purposes were typically much closer to the purposes of states. The charters of these early corporations stipulated that the corporation ought to further the common good. They were assigned for only 20 years and withdrawn if the common good wasn't furthered.

⁵⁴Alchian & Demsetz (1972, 779)

hard than she agreed to, leading to *free-rider problems and shirking*.⁵⁵ These costs for team productions also emerge whenever the allocation of benefits is not fixed from the start. This gives rise to situations in which each member in the joint enterprise engages in *rent-seeking* behaviour, trying to maximise her share of the corporate surplus.

To address the costs of shirking and rent-seeking which are characteristic for team productions, it is not an efficient solution, as Stout & Blair (1999) argue, to simply introduce a principal who owns the corporate assets of the corporation, punishes shirking and rent-seeking behaviour and is compensated for that task with the profits of the joint enterprise. Such a solution is itself threatened by the principal's own potential rent-seeking behaviour. Rather, the best solution emerges, when each team member makes an irrevocable commitment of their resources, their firm-specific investments, to the joint enterprise, thereby giving up control rights over their investment to the joint enterprise.⁵⁶ Unlike the principal-agent form, where the assets are held by the principal(s), this solution implies that the assets are held by the firm *itself*. On this approach, there is no principal in whose interests the enterprise can be governed, the firm must instead be represented by a third party that has itself made no firm-specific investment. Otherwise, that party will use the control it has gained to unilaterally increase its rents over the enterprise at the expense of the other team members. This third party is typically known, in public corporations, as the 'board of directors', which has the final say over the company's directions. The resulting hierarchy is called a *mediating hierarchy model*, since the directors, who represent the company, mediate between the different interests of the groups that which have a stake in the production, without being the sole arbiter of the shareholders or any other specific stakeholder group.⁵⁷

⁵⁵These problems also appear under the heading of the prisoner's dilemma or social dilemmas more generally. They have given rise to an enormous literature across a wide range of fields, from game theory to philosophy and the social sciences.

⁵⁶Funnily enough, this is exactly similar to the way Margaret Gilbert's describes the formation of plural subjects.

⁵⁷There is an expansive literature on how the kind of service a corporation offers can influence which organisational structure best solves the underlying cooperation problems. Hansmann (1980), for example, has argued that highly complex products such as services offered by hospitals or services offered by a University can lead to specific kinds of contracting costs that are most efficiently resolved by non-profit corporations characterised by a 'non-distribution' constraint and which forbids the managers from pocketing the profits that their organisation generates. Hansmann argues that such a constraint 'forces' the managers to reinvest the profits back into the organisation rather than extract them for personal use.

The Normative Profile of the Stakeholder Governed Corporation

A corporation with a point of view that is shaped by the StG-rule has greater leeway in choosing its ends. Unlike the ODV, it not only recognizes morally salient facts, but can, through adapting its web of shared policies of giving weight, respond in a way that is required by morality. The StG-rule will have already been interpreted in some way, depending on previous choices, commitments, contextual facts that make certain interpretation more likely than others. However, the possibility to adjust the weights of the shared policies of giving weights, given that there are no maximising singular constitutive ends, can provide those who must choose on behalf of the corporation with a feeling of epistemic freedom that ensures the kind of self-intelligibility required of fully autonomous actions.

But, typically, things will not be as straight forward. Even if a public corporation sees itself as being governed by the StG-rule and is seen in this way by others, certain facts that give rise to the arguments that have been forwarded for shareholder governance, like competitive markets, legal pressure, shareholder lobbying can lead to an interpretation of the StG-rule that reduces it to the value maximization view supported by Jensen. If these circumstantial factors prove systematic, then even a potentially moral agent might be excused for a period of time which might amount to an exemption similar in spirit as the one the ODV is under.

6.5 Summing up

In the chapters leading up to this, I have argued that if certain groups have a certain kind of end, then they are under rational pressure to realize a functional profile of an ODV that undermines their autonomous agency. I have also proposed a constitutive account of agency, which is supposed to capture the janus facedness of groups as means and as, in principle, capable of autonomous action. What I promised to do, was compile empirical evidence that groups with a functional structure like the ODV exist. I hope to have made it plausible that the debate on shareholder governance can and has been interpreted in a way that mirrors the normative profile of the ODV. I have also argued that this debate not only presupposes a social practices that includes different various social entities, but that in certain social environments it is plausible to suppose that this practice exists and shapes the group agents that find itself encumbered by it.

Importantly not every practice that is captured by the umbrella term shareholder governance is such that it puts rational pressure on group agents

to become ODVs. These and social practices that fall under stakeholder governance might allow groups to act autonomously.

Conclusion

Some philosophers take it to be obvious that corporations are moral agents, and that they can be held responsible in a way similar to how we hold each other responsible. Although I was convinced by arguments establishing the possibility of rational agency by groups like corporations, the seamless transition to the case for group moral agency by philosophers like List and Pettit, for example, seemed to omit certain characteristic features of groups. It left out an important aspect which confounds business ethicists interested in the aims of corporations: squaring a corporation's perceived function with, depending on the kind of market, sometimes incompatible moral demands.

Perhaps, from the perspective of ideal theory, this struggle is simply confused, boiling down to Korsgaard's normative question whether 'moral requirements provide genuine reasons for acting regardless of an agent's contingent desires and interests.'⁵⁸ Although this is a possibility, the problem seems to have a different kind of urgency in the case of groups which lack the motivating force that we experience when feeling reactive emotions.⁵⁹ But thinking of this conundrum for business ethicists from a purely normative standpoint gets it wrong. Normatively justifying one or another institutional aim of, say, the public corporation is only part of their goal in the debate on corporate governance. Rather, the struggle seems to be how to reconcile the social fact that people simply believe that – or, as I argue, participate in a convention according to which – the public corporation is simply assigned a certain end and the demands of morality. As I have argued, this conflict involves the idiosyncratic agency of corporations, but cannot be captured by the existing theories of group agency.

The constitutive account of group agency asserts that what is constitutive of autonomous group action is being governed by group self-intelligibility.

⁵⁸(FitzPatrick 2005, 653)

⁵⁹For this argument, see Sepinwall & Orts (2017). Although a number of philosophers, such as Tollefsen (2002), Björnsson & Hess (2017) and Hindriks (2018), have argued that groups such as corporations can possess a functional equivalent of these emotions, I remain unconvinced for reasons that groups can have normative profiles that are categorically different to individual agents.

This, combined with the functionalist account of group agency, was my attempt at providing a framework which captures both faces of group agency in a unifying way. I believe that it provides us with novel ways of exploring and investigating the difference between agency in institutional group agents such as corporations, NGO's, football clubs and churches and individual agency. Simultaneously, it allows us to harness rich insights from the philosophy of action. The difference, I have claimed, lies in how a group's practical identity is grounded compared to a human agent's practical identity. In most groups, that identity is grounded in an externally imposed function that is collectively accepted by the relevant parts of society and which is, therefore, unilaterally unalterable from the group's perspective. The individual, on the other hand, can always fall back to their identity as a human being with its associated dignity, and make sense of moral issues from that perspective.

I want to conclude this dissertation with a few thoughts about the example I began with: Samsung's letter to the American public. In this letter, Samsung expresses guilt for having fallen short of their promise and their responsibility for consumer safety and quality after reports on exploding Galaxy phones appeared. The letter was clearly meant to convey the image of Samsung as a caring and responsible agent, an agent whom one can trust and who deserves forgiveness. The case is interesting because Samsung seems to actively endorse the idea that it is an agent sensitive to considerations of morality and value. As such, it is more subtle than certain recent cases of psychopathic behaviour by pharmaceutical companies or fossil fuel corporations that fight against climate science. Against this, the constitutive account of group agency justifies the attribution of another functional and normative profile along the lines of a one-dimensional valuing agent governed by the DSG rule explored in the final chapter.

Samsung, a public corporation, operates in a hyper-competitive, global market environment. In it, corporations are subject to enormous pressure to stay ahead of the competition by constantly developing new smartphones, even if this means pushing the limits of technological possibility. Together, these facts suggests that Samsung's practical identity consists of the DSG rule. Several newspaper articles have suggested that the practice that this rule gives rise to was responsible for the risky design choice that turned the Galaxy phone into a safety hazard.⁶⁰ Several decisions make plausible the idea that the aim of maximizing profits by minimizing losses was shaping the company's deliberation and decision making. Although they voluntarily offered a recall of the phone, some interpreted this offer as a way to avoid a situation that Toyota was facing at the time, related to false accelera-

⁶⁰ *5 lessons to be learnt from Samsung's Note 7 fiasco* (n.d.)

tion. Government agencies forced Toyota to issue the recall, which resulted in expensive governmental fines and lawsuits.⁶¹ Furthermore, although the decision to issue a recall of the phone in the United States came after two months,⁶² it was delayed in China. Some attributed this delay to the size of the Chinese market.⁶³ With regard to other, lower-profile products like washing machines, Samsung's recall policy was judged to be slow and bureaucratic. This seems to imply that non-flagship products that had no potential to damage the brand's value did not receive the same consistent treatment.⁶⁴

Of course, this case does not, by itself, prove my point that Samsung is a one-dimensional valuer. At best, it shows that the company's actions are compatible with such a functional profile. But it does caution us against rushed ascriptions of moral agency. Even if it turns out that public corporations like Samsung are one-dimensional valuers, if they care enough about the value connected to their reputation,⁶⁵ we might have enough leverage on them within merely aretaic responsibility practices. But not every public company might care about their reputation, so we should try as much as possible to contain their power and keep a watchful eye on tool-like group agents that are turned against their makers.

⁶¹*Legal Implications of Samsung's Huge Recall of Galaxy Note 7 Phones* (n.d.)

⁶²*TIMELINE-Samsung Electronics' Galaxy Note 7 recall crisis* (n.d.)

⁶³*The Samsung Galaxy Note 7 Recall Spreads Late to China* (n.d.)

⁶⁴*Galaxy Note 7 Is Not Samsung's Only Problematic Product* (n.d.)

⁶⁵*Survey Suggests Galaxy Note 7 Recall Did Not Damage Samsung Brand* (n.d.)

Bibliography

- 5 lessons to be learnt from Samsung's Note 7 fiasco (n.d.), <https://www.businesstoday.in/technology/news/story/5-lessons-to-be-learnt-from-samsungs-note-7-fiasco-68835-2016-10-13>. Accessed: 2022-03-01.
- Albertzart, M. (2017), Monsters and their makers: Group agency without moral agency, in 'Reflections on ethics and responsibility', Springer, pp. 21–35.
- Alchian, A. A. & Demsetz, H. (1972), 'Production, information costs, and economic organization', *The American economic review* **62**(5), 777–795.
- Alonso, F. M. (2009), 'Shared intention, reliance, and interpersonal obligations', *Ethics* **119**(3), 444–475.
- Alonso, F. M. (2016), 'A dual aspect theory of shared intention', *Journal of Social Ontology* **2**(2), 271–302.
- Altner, F. (2021), 'A counterfactual account of diachronic structural rationality', *Inquiry* pp. 1–30.
- Anscombe, G. E. M. (2000), *Intention*, Harvard University Press.
- Aumann, R. J. (1997), Rationality and bounded rationality, in 'Cooperation: Game-Theoretic Approaches', Springer, pp. 219–231.
- Austin, J. L. (1971), A plea for excuses, in 'Philosophy and linguistics', Springer, pp. 79–101.
- Bacharach, M., Gold, N. & Sugden, R. (2006), *Beyond individual choice: teams and frames in game theory*, Princeton University Press.
- Baumol, W. J. & Blackman, S. A. B. (1991), 'Perfect markets and easy virtue business ethics and the invisible hand'.

-
- Berle, A. A. (1932), 'For whom corporate managers are trustees: A note', *Harvard law review* **45**(8), 1365–1372.
- Berle, A. A. & Means, G. G. C. (1932), *The modern corporation and private property*, New York: Macmillan.
- Björnsson, G. & Hess, K. (2017), 'Corporate crocodile tears? on the reactive attitudes of corporate agents', *Philosophy and Phenomenological Research* **94**(2).
- Block, N. (1980), 'Troubles with functionalism', *Readings in philosophy of psychology* **1**, 268–305.
- Boatright, J. R. (1994), 'Fiduciary duties and the shareholder-management relation: Or, what's so special about shareholders?', *Business Ethics Quarterly* pp. 393–407.
- Bratman, M. (1987), *Intention, Plans, and Practical Reason*, Center for the Study of Language and Information.
- Bratman, M. (1999), *Faces of intention: Selected essays on intention and agency*, Cambridge University Press.
- Bratman, M. E. (1992), 'Shared cooperative activity', *The philosophical review* **101**(2), 327–341.
- Bratman, M. E. (1993), 'Shared intention', *Ethics* **104**(1), 97–113.
- Bratman, M. E. (2000), 'Valuing and the will', *Noûs* **34**, 249–265.
- Bratman, M. E. (2012), 'Time, rationality, and self-governance', *Philosophical Issues* **22**(1), 73–88.
- Bratman, M. E. (2013), *Shared agency: A planning theory of acting together*, Oxford University Press.
- Bratman, M. E. (2017), 'The intentions of a group', *The moral responsibility of firms* pp. 36–52.
- Bratman, M. E. (2018), *Planning, time, and self-governance: essays in practical rationality*, Oxford University Press.
- Broome, J. (2013), *Rationality through reasoning*, John Wiley & Sons.
- Chalmers, D. J. (1996), *The conscious mind: In search of a fundamental theory*, Oxford Paperbacks.

-
- Chiao, V. (2014), 'List and Pettit on group agency and group responsibility', *University of Toronto Law Journal* **64**(5), 753–770.
- Clark, R. C. (1983), *Agency costs versus fiduciary duties*, Division of Research, Harvard Business School.
- Coase, R. H. (1960), The problem of social cost, in 'Classic papers in natural resource economics', Springer, pp. 87–137.
- Coffee, J. C. (1981), "'No soul to damn: no body to kick": An unsandalized inquiry into the problem of corporate punishment', *Michigan Law Review* **79**(3), 386–459.
- Collins, S. (2019), *Group duties: Their existence and their implications for individuals*, Oxford University Press.
- Collins, S. & Lawford-Smith, H. (2016), 'Collectives' and individuals' obligations: a parity argument', *Canadian Journal of Philosophy* **46**(1), 38–58.
- Davidson, D. (1985), 'Incoherence and irrationality', *dialectica* **39**(4), 345–354.
- de Haan, N. (2022), 'Moral agency as a catch 22', *under review*.
- Dennett, D. C. (1981), 'True believers: The intentional strategy and why it works'.
- Dennett, D. C. (1989), *The intentional stance*, MIT press.
- Dodd Jr, E. M. (1931), 'For whom are corporate managers trustees', *Harv. L. Rev.* **45**, 1145.
- Drucker, P. F. (1984), 'Converting social problems into business opportunities: The new meaning of corporate social responsibility', *California Management Review (pre-1986)* **26**(000002), 53.
- Eisenberg, M. A. (1998), 'The conception that the corporation is a nexus of contracts, and the dual nature of the firm', *J. Corp. L.* **24**, 819.
- Epstein, B. (2015), *The ant trap: Rebuilding the foundations of the social sciences*, Oxford Studies in Philosophy.
- Epstein, B. (2019), 'What are social groups? their metaphysics and how to classify them', *Synthese* **196**(12), 4899–4932.

Eshleman, A. (2016), Moral responsibility, in E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', winter 2016 edn, Metaphysics Research Lab, Stanford University.

Exxon loses board seats to activist hedge fund in landmark climate vote (n.d.), <https://www.reuters.com/business/sustainable-business/shareholder-activism-reaches-milestone-exxon-board-vote-nears-end-2021-05-26/>. Accessed: 2021-06-30.

Ferrero, L. (2010), 'Decisions, diachronic autonomy, and the division of deliberative labor', *Philosophers* **10**.

Ferrero, L. (2014), 'Diachronic structural rationality', *Inquiry* **57**(3), 311–336.

Fischer, J. M. (2004), 'Responsibility and manipulation', *The Journal of Ethics* pp. 145–177.

Fischer, J. M. (2012), 'Responsibility and autonomy: The problem of mission creep', *Philosophical Issues* **22**, 165–184.

FitzPatrick, W. J. (2005), 'The practical turn in ethical theory: Korsgaard's constructivism, realism, and the nature of normativity', *Ethics* **115**(4), 651–691.

Fodor, J. A. (1975), *The language of thought*, Vol. 5, Harvard university press.

Frankfurt, H. (2018), Freedom of the will and the concept of a person, in 'Agency And Responsibility', Routledge, pp. 77–91.

Frankfurt, H. B. & Frankfurt, H. G. (1999), *Necessity, volition, and love*, Cambridge University Press.

Frankfurt, H. G. (2006), *Taking ourselves seriously and getting it right*, Stanford University Press.

French, P. A. (1979), 'The corporation as a moral person', *American Philosophical Quarterly* **16**(3), 207–215.

French, P. A. (1984), *Collective and corporate responsibility*, Columbia University Press.

Friedman, M. (2007), The social responsibility of business is to increase its profits, in 'Corporate ethics and corporate governance', Springer, pp. 173–178.

- Galaxy Note 7 Is Not Samsung's Only Problematic Product* (n.d.), <https://www.nytimes.com/2016/10/13/business/international/samsung-galaxy-note7-profit-battery-fires.html?searchResultPosition=5>. Accessed: 2022-03-01.
- Gauthier, D. (1986), *Morals by agreement*, Oxford University Press on Demand.
- Gauthier, D. (1994), 'Assure and threaten', *Ethics* **104**(4), 690–721.
- Gauthier, D. (1998), 'Intention and deliberation', *Modeling Rationality, Morality, and Evolution* pp. 41–54.
- Gigerenzer, G. & Goldstein, D. G. (1996), 'Reasoning the fast and frugal way: models of bounded rationality.', *Psychological review* **103**(4), 650.
- Gilbert, M. (1987), 'Modelling collective belief', *Synthese* **73**(1), 185–204.
- Gilbert, M. (1990), 'Walking together: A paradigmatic social phenomenon', *MidWest studies in philosophy* **15**(1), 1–14.
- Gilbert, M. (1992), *On social facts*, Princeton University Press.
- Gilbert, M. (2006), *A theory of political obligation: Membership, commitment, and the bonds of society*, OUP Oxford.
- Gold, N. & Sugden, R. (2007), *Theories of team agency*, Oxford University Press.
- Goodpaster, K. E. (1991), 'Business ethics and stakeholder analysis', *Business ethics quarterly* pp. 53–73.
- Greenspan, P. (2016), 'Responsible psychopaths revisited', *The Journal of Ethics* **20**(1-3), 265–278.
- Greenspan, P. S. (2003), 'Responsible psychopaths', *Philosophical Psychology* **16**(3), 417–429.
- Grice, H. P. (1975), Logic and conversation, in 'Speech acts', Brill, pp. 41–58.
- Grice, P. (2013), 'Method in philosophical psychology (from the banal to the bizarre)', *The American Philosophical Association Centennial Series* pp. 305–332.
- Hansmann, H. B. (1980), 'The role of nonprofit enterprise', *The Yale law journal* **89**(5), 835–901.

-
- Harman, G. (1986), *Change in view: Principles of reasoning*, The MIT Press.
- Haslanger, S. (2016), 'What is a (social) structural explanation?', *Philosophical Studies* **173**(1), 113–130.
- Hasnas, J. (1998), 'The normative theories of business ethics: A guide for the perplexed', *Business Ethics Quarterly* **8**(1), 19–42.
- Hausman, D. M. (1994), *The philosophy of economics: An anthology*, Cambridge University Press.
- Heath, J. (2014), *Morality, competition, and the firm: The market failures approach to business ethics*, Oxford University Press.
- Heath, J. & Norman, W. (2004), 'Stakeholder theory, corporate governance and public management: what can the history of state-run enterprises teach us in the post-enron era?', *Journal of business ethics* **53**(3), 247–265.
- Hess, K. M. (2010), 'The modern corporation as moral agent: The capacity for "thought" and a "first-person perspective"', *Southwest Philosophy Review* **26**(1), 61–69.
- Hess, K. M. (2012), 'Group agency: The possibility, design, and status of corporate agents—by christian list & philip pettit', *Journal of Applied Philosophy* **29**(2), 165–167.
- Hess, K. M. (2014), 'The free will of corporations (and other collectives)', *Philosophical Studies* **168**(1), 241–260.
- Hindriks, F. (2008), 'The status account of corporate agents', *Concepts of sharedness - essays on collective intentionality* pp. 119–144.
- Hindriks, F. (2014), 'How autonomous are collective agents? corporate rights and normative individualism', *Erkenntnis* **79**(9), 1565–1585.
- Hindriks, F. (2018), 'Collective agency: Moral and amoral', *Dialectica* **72**(1), 3–23.
- Hindriks, F. A. (2003), 'The new role of the constitutive rule', *American Journal of Economics and Sociology* **62**(1), 185–208.
- Hindriks, F. & Guala, F. (2015), 'Institutions, rules, and equilibria: a unified theory', *Journal of Institutional Economics* **11**(3), 459–480.

-
- Hindriks, F. & Guala, F. (2019), 'The functions of institutions: etiology and teleology', *Synthese* pp. 1–17.
- Holton, R. (2004), 'Rational resolve', *The Philosophical Review* **113**(4), 507–535.
- Horowitz, S. (2014), 'Epistemic akrasia', *Noûs* **48**(4), 718–744.
- Hume, D. et al. (2000), *An enquiry concerning human understanding: A critical edition*, Vol. 3, Oxford University Press.
- Hussain, W. (2012), 'Corporations, profit maximization and the personal sphere', *Economics & Philosophy* **28**(3), 311–331.
- Jensen, M. C. (2001), 'Value maximization, stakeholder theory, and the corporate objective function', *Journal of applied corporate finance* **14**(3), 8–21.
- Jensen, M. C. & Meckling, W. H. (1976), 'Theory of the firm: Managerial behavior, agency costs and ownership structure', *Journal of financial economics* **3**(4), 305–360.
- Kahneman, D. (2011), *Thinking, fast and slow*, Macmillan.
- Kahneman, D. & Tversky, A. (2013), Choices, values, and frames, in 'Handbook of the fundamentals of financial decision making: Part I', World Scientific, pp. 269–278.
- Kiesewetter, B. (2017), *The normativity of rationality*, Oxford University Press.
- Kolodny, N. (2007), IX—how does coherence matter?, in 'Proceedings of the Aristotelian Society (Hardback)', Vol. 107, Wiley Online Library, pp. 229–263.
- Kolodny, N. (2008), 'Why be disposed to be coherent?*', *Ethics* **118**, 437–463.
- Korsgaard, C. M. (1996), *The sources of normativity*, Cambridge University Press.
- Korsgaard, C. M. (2003), 'Realism and constructivism in twentieth-century moral philosophy', *Journal of Philosophical Research* **28**(Supplement), 99–122.
- Kraakman, R. & Hansmann, H. (2017), The end of history for corporate law, in 'Corporate Governance', Gower, pp. 49–78.

-
- Kusch, M. (2014), 'The metaphysics and politics of corporate personhood', *Erkenntnis* **79**(9), 1587–1600.
- Legal Implications of Samsung's Huge Recall of Galaxy Note 7 Phones* (n.d.), <https://lawandcrime.com/high-profile/legal-implications-of-samsungs-huge-recall-of-galaxy-note-7-phones/>. Accessed: 2022-03-01.
- Leggett, C. (1999), 'The Ford Pinto case: The valuation of life as it applies to the negligence-efficiency argument', *Law and Valuation Papers*.
- Lewis, D. (2008), *Convention: A philosophical study*, John Wiley & Sons.
- List, C. (2012), 'The theory of judgment aggregation: An introductory review', *Synthese* **187**(1), 179–207.
- List, C. & Pettit, P. (2002), 'Aggregating sets of judgments: An impossibility result', *Economics & Philosophy* **18**(1), 89–110.
- List, C. & Pettit, P. (2011), *Group agency: The possibility, design, and status of corporate agents*, Oxford University Press.
- Lord, E. (2018), *The importance of being rational*, Oxford University Press.
- Ludwig, K. (2017a), 'Do corporations have minds of their own?', *Philosophical Psychology* **30**(3), 269–301.
- Ludwig, K. (2017b), *From Plural to Institutional Agency: Collective Action II*, Oxford University Press.
- Macey, J. R. (1991), 'An economic analysis of the various rationales for making shareholders the exclusive beneficiaries of corporate fiduciary duties', *Stetson L. Rev.* **21**, 23.
- MacIntyre, A. (2013), *After virtue*, A&C Black.
- Mankiw, N. G. (2014), *Principles of economics*, Cengage Learning.
- McKenna, M. (2014), 'Resisting the manipulation argument: A hard-liner takes it on the chin', *Philosophy and Phenomenological Research* **89**(2), 467–484.
- Mele, A. R. (2005), 'A critique of Pereboom's 'four-case argument' for incompatibilism', *Analysis* **65**(1), 75–80.

-
- Miller, S. (2003), Social institutions, in 'Realism in Action', Springer, pp. 233–249.
- Miller, S. (2006), 'Collective moral responsibility: An individualist account', *Midwest studies in philosophy* **30**(1), 176–193.
- Miller, S. et al. (2001), *Social action: A teleological account*, Cambridge University Press.
- Millikan, R. G. (2000), 'Reading mother nature's mind', *Dennett's philosophy: A comprehensive assessment* pp. 55–75.
- Moriarty, J. (2017), Business ethics, in E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', fall 2017 edn, Metaphysics Research Lab, Stanford University.
- Ostrander, R. C. (Mich. 1919), 'Dodge v. Ford Motor Co.', **170 N.W. 668**.
- Pacherie, E. (2011), 'Framing joint action', *Review of Philosophy and Psychology* **2**(2), 173–192.
- Pauer-Studer, H. (2014), 'A constitutive account of group agency', *Erkenntnis* **79**(9), 1623–1639.
- Pearce, J. M. (2019), 'Towards quantifiable metrics warranting industry-wide corporate death penalties', *Social Sciences* **8**(2), 62.
- Pereboom, D. (2006), *Living without free will*, Cambridge University Press.
- Petersson, B. (2007), 'Collectivity and circularity', *The Journal of Philosophy* **104**(3), 138–156.
- Pettit, P. (2003), 'Akrasia, collective and individual', *Weakness of will and practical irrationality* pp. 68–96.
- Pettit, P. (2017), 'The conversable, responsible corporation', *The moral responsibility of firms* pp. 15–33.
- Poprawe, F. (2020), 'Is reasoning responding to reasons?', *Philosophical Explorations* **23**(2), 146–159.
- Quinn, D. P. & Jones, T. M. (1995), 'An agent morality view of business policy', *Academy of Management Review* **20**(1), 22–42.
- Rachar, M. (2021), 'Quasi-psychologism about collective intention', *Ethical Theory and Moral Practice* pp. 1–14.

-
- Rawls, J. (1955), 'Two concepts of rules', *The philosophical review* **64**(1), 3–32.
- Raz, J. (2005), 'The myth of instrumental rationality', *J. Ethics & Soc. Phil.* **1**, 1.
- Rönnegard, D. (2015), *The Fallacy of Corporate Moral Agency*, Vol. 44, Springer.
- Rovane, C. (1997), *The bounds of agency: An essay in revisionary metaphysics*, Princeton University Press.
- Samsung runs full-page apology ads over Galaxy Note 7 recall* (n.d.), <https://www.theverge.com/2016/11/7/13558442/samsung-galaxy-note-7-recall-apology-ad>. Accessed: 2022-03-01.
- Scanlon, T. et al. (1998), *What we owe to each other*, Harvard University Press.
- Schmid, H. B. (2009), *Plural action: Essays in philosophy and social science*, Vol. 58, Springer Science & Business Media.
- Schmid, H. B. (2014), 'Plural self-awareness', *Phenomenology and the Cognitive Sciences* **13**(1), 7–24.
- Schumpeter (July 6th 2019), 'Ben van Beurden's balancing act', *The Economist* **9150**, 55.
- Searle, J. R. (1990), 'Collective intentions and actions', *Intentions in communication* **401**, 401–02.
- Searle, J. R. (1992), *The rediscovery of the mind*, MIT press.
- Searle, J. R. (2005), 'What is an institution?', *Journal of institutional economics* **1**(1), 1–22.
- Searle, J. R. (2006), 'Social ontology: Some basic principles', *Anthropological theory* **6**(1), 12–29.
- Searle, J. R., Willis, S. et al. (1983), *Intentionality: An essay in the philosophy of mind*, Cambridge university press.
- Searle, J. R., Willis, S. et al. (1995), *The construction of social reality*, Simon and Schuster.

-
- Sen, A. (1985), 'The moral standing of the market', *Social philosophy and policy* **2**(2), 1–19.
- Sepinwall, A. J. & Orts, E. W. (2017), 'Blame, emotion, and the corporation', *The moral responsibility of firms* pp. 143–166.
- Shafer-Landau, R. (2012), *Ethical theory: an anthology*, Vol. 13, John Wiley & Sons.
- Shapiro, S. (2014), 'Massively shared agency', *Rational and social agency: The philosophy of Michael Bratman* pp. 257–293.
- Shoemaker, D. (2011), 'Attributability, answerability, and accountability: Toward a wider theory of moral responsibility', *Ethics* **121**(3), 602–632.
- Shoemaker, S. (1996), *The first-person perspective and other essays*, Cambridge University Press.
- Silver, K. (2021), 'Group action without group minds', *Philosophy and Phenomenological Research* .
- Silverstein, M. (2015), 'The shmagency question', *Philosophical Studies* **172**(5), 1127–1142.
- Soderquist, L. D. & Vecchio, R. P. (1978), 'Reconciling shareholders' rights and corporate responsibility: New guidelines for management', *Duke LJ* p. 819.
- Stout, L. A. (2001), 'Bad and not-so-bad arguments for shareholder primacy', *S. Cal. L. Rev.* **75**, 1189.
- Stout, L. A. (2008), 'Why we should stop teaching dodge v. ford', *Va. L. & Bus. Rev.* **3**, 163.
- Stout, L. A. & Blair, M. M. (1999), A team production theory of corporate law, in 'Virginia Law Review Association', Vol. 85, HeinOnline, pp. 247–328.
- Strawson, P. F. (2008), *Freedom and resentment and other essays*, Routledge.
- Structure and Governance - Website of Proctor and Gamble* (n.d.), <https://us.pg.com/structure-and-governance/corporate-governance-overview/>. Accessed: 2022-03-14.
- Sugden, R. (2000), 'Team preferences', *Economics & Philosophy* **16**(2), 175–204.

- Sugden, R. (2012), 'Must group agents be rational? list and pettit's theory of judgement aggregation and group agency', *Economics & Philosophy* **28**(2), 265–273.
- Survey Suggests Galaxy Note 7 Recall Did Not Damage Samsung Brand* (n.d.), <https://fortune.com/2016/11/21/galaxy-note-7-recall-samsung-brand/>. Accessed: 2022-03-01.
- Szanto, T. (2016), Do group persons have emotions - or should they?, in 'Analytic and Continental Philosophy: Methods and Perspectives. Proceedings of the 37th International Wittgenstein Symposium', Vol. 23, Walter de Gruyter GmbH & Co KG, p. 261.
- Tenenbaum, S. (2018), 'Reconsidering intentions', *Noûs* **52**(2), 443–472.
- The Samsung Galaxy Note 7 Recall Spreads Late to China* (n.d.), <https://fortune.com/2016/09/16/samsung-galaxy-recall-china/>. Accessed: 2022-03-01.
- TIMELINE-Samsung Electronics' Galaxy Note 7 recall crisis* (n.d.), <https://www.reuters.com/article/samsung-elec-smartphones-idUSL4N1CG1KB>. Accessed: 2022-03-01.
- Tollefsen, D. (2002), 'Organizations as true believers', *Journal of social philosophy* **33**(3), 395–410.
- Tollefsen, D. (2005), 'Let's pretend! children and joint action', *Philosophy of the Social Sciences* **35**(1), 75–97.
- Tubert, A. (2010), 'Constitutive arguments', *Philosophy Compass* **5**(8), 656–666.
- Tuomela, R. (1995), 'The importance of us: A philosophical study of basic social notions'.
- Tuomela, R. (2013), *Social ontology: Collective intentionality and group agents*, Oxford University Press.
- Vaesen, K. & Van Amerongen, M. (2008), 'Optimality vs. intent: limitations of dennett's artifact hermeneutics', *Philosophical Psychology* **21**(6), 779–797.
- Van Fraassen, B. C. (1973), 'Values and the heart's command', *The Journal of Philosophy* **70**(1), 5–19.

-
- Velleman, J. D. (2006), *Self to self: Selected essays*, Cambridge University Press.
- Velleman, J. D. (2009a), *How we get along*, Cambridge University Press.
- Velleman, J. D. (2009b), *The possibility of practical reason*, Michigan Publishing, University of Michigan Library.
- Velleman, J. D. (2012), What good is a will?, in 'Action in context', de Gruyter, pp. 193–215.
- Vermaas, P. E., Carrara, M., Borgo, S. & Garbacz, P. (2013), 'The design stance and its artefacts', *Synthese* **190**(6), 1131–1152.
- Wallace, R. J. (1994), *Responsibility and the moral sentiments*, Harvard University Press.
- Watson, G. (1987), 'Free action and free will', *Mind* **96**(382), 145–172.
- Watson, G. (1996), 'Two faces of responsibility', *Philosophical Topics* **24**(2), 227–248.
- Williams, B. (1981), *Moral luck: philosophical papers 1973-1980*, Cambridge University Press.
- Williams, B. (2008), *Shame and necessity*, University of California Press.
- Williamson, O. E. (2007), The economic institutions of capitalism. firms, markets, relational contracting, in 'Das Summa Summarum des Management', Springer, pp. 61–75.
- Wolf, S. (1987), 'Sanity and the metaphysics of responsibility', 1987 pp. 46–62.
- Wolf, S. (1993), *Freedom within reason*, Oxford University Press on Demand.
- Wringe, B. (2014), 'May i treat a collective as a mere means?', *American Philosophical Quarterly* **51**(3), 273–284.