Group Character

A Study of Collective Virtue and Vice

PhD thesis

to obtain the degree of PhD of the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijminga
and in accordance with the decision
by the College of Deans

and

to obtain the degree of PhD of the
University of Sheffield
on the authority of the
President and Vice-Chancellor Koen Lamberts
and in accordance with the decision
by the Faculty of Arts and Humanities

Joint PhD degree

This thesis will be defended in public on

Thursday 13 October 2022 at 16:15 hours

by

Barend Cornelis de Rooij

born on 20 March 1992
in Tilburg

Supervisors
Prof. B. P. de Bruin
Prof. M. Fricker

Assessment Committee
Prof. H. Battaly
Prof. L. Herzog
Prof. G. J. de Ridder
Prof. R. Stern

Table of Contents

Acknowledgements

In the course of writing this dissertation on group character and jointly intentional activity, it has become abundantly clear that completing a Ph.D. is a profoundly social enterprise itself. Accordingly, I have been fortunate enough to experience a recurring theme in the literature on virtue ethics firsthand: flourishing in *any* social activity crucially depends on the support, wisdom, and goodwill of others. It is not surprising, then, that I owe a deep debt of gratitude to the many people who created the conditions enabling this research project to thrive.

First of all, I owe warmest thanks to my primary supervisors, Boudewijn de Bruin and Miranda Fricker. I am extremely grateful for the opportunity to be a part of the *Towards Professional Epistemic Justice* project and work alongside such talented philosophers. You not only ensured that this project was off to a great start, but also kept it on track throughout. Managing a collaborative international research project cannot have been easy in the midst of a global pandemic, but despite all hardships you did an excellent job of making everyone involved stay connected. Thank you both for your support and encouragement.

Boudewijn, I admire your philosophical rigor and your passion for applying cutting edge philosophical research to current affairs. Your detailed feedback on the many drafts of this dissertation has greatly contributed to the final product before you today. Thank you, also, for offering to collaborate with me on the Boeing chapter, and for the many other doors you have opened for me. Besides a great philosopher, you are a fantastic manager, never forgetting the people behind the work. These are lessons I will always carry with me.

Miranda, your sharp philosophical insight and keen awareness of how philosophical problems pertain to contemporary social issues have inspired me from day one. The relevance of the field you pioneered fifteen years ago continues to prove itself every day, and I am deeply grateful for the opportunity to explore topics of justice, epistemology, and virtue under your supervision. Whenever I tangled myself (or my arguments) in a knot, you were invaluable in untangling them and helping me see the bigger picture. Thank you so much for all of your feedback and assistance during the last four years.

During my two years at Sheffield University, I had the privilege of working with Paul Faulkner as my secondary supervisor. Paul, I am very grateful for your support, and for

detailed comments on chapters one and five of this dissertation. The brainstorming sessions we had together were incredibly fruitful, and your critical eye has been instrumental in sharpening the focus of my writing. Thank you also for being my main point of contact in Sheffield after Miranda's move to New York.

As important as supervisors are, a Ph.D. cannot be completed unless it is approved by the assessment committee. Thank you, Heather Battaly, Lisa Herzog, Jeroen de Ridder, and Bob Stern, for taking the time to read this dissertation, and for subjecting it to rigorous scrutiny during my defense. If you found it even half as stimulating as I find your research, I would be very happy. Heather, I am particularly indebted to your thinking about (epistemic) virtue and vice – thank you for bringing clarity to such a complex, rapidly evolving field.

Jeroen, I owe thanks to you not only in your capacity as member of the assessment committee, but also for your detailed comments on chapter four of this dissertation and your editorial support in getting that chapter published. I would also like to thank the other editors of *Social Virtue Epistemology*, Mark Alfano and Colin Klein. Further thanks goes out to Steven Bland, Frank Hindriks, and Neil Levy, who also provided comments on chapter four.

The *Towards Professional Epistemic Justice* project is much bigger than the contributions of Boudewijn, Miranda, and myself alone. As such, I would also like to extend my gratitude to my colleagues, Kayleigh Doherty and Elianna Fetterolf. Kayleigh, you are not only a great philosopher, but have also become a dear friend. Elianna, thank you for making an impact by applying research on epistemic justice in actual healthcare settings. In addition, I want to acknowledge the generous support of the institutions that helped this project take off: the NWO, who provided the funding, and the University of Sheffield and Rijksuniversiteit Groningen, who provided a stimulating research environment.

One of the great perks of doing academic research is that you get to travel the world to talk about your studies with likeminded people. Many thanks to audiences in Birmingham, Rijeka, Sheffield, Tampere, Vienna, and Tilburg for offering your thoughts on parts of this dissertation. Ideas can sound really good in your head until you test them out in front of others. Thank you all for showing me the way forward.

The last four years have not just been a full of academic discoveries, but also personal ones. I am especially grateful to have discovered a deep passion for teaching. As such, I would like to thank my colleagues at Groningen's Faculty of Economics and Business for giving me the opportunity to teach business ethics. I cannot express in words how much I have learned from the example of Chiara Lisciandra, Marijke Leliveld, Juliette de Wit, Wim Westerman, and Raymond Zaal. Further thanks goes out to the excellent instruction on the art

Introduction

When U.S. presidential candidate Hillary Clinton referred to her opponent Donald Trump's supporters as a "basket of deplorables" during a 2016 campaign speech, she was fiercely criticized for writing off millions of Americans.[1] Clarifying her remark, Clinton responded that her issue was not with the character of individual Trump supporters, but with the hateful, divisive views embodied by Trump's movement as a whole. Though Clinton's riposte was unable to assuage all concerns about her views of individual Trump voters, her rebuttal is not entirely without *philosophical* merit. In suggesting that Trump's movement is capable of expressing views and attitudes that do not reflect the character of its individual members, Clinton's response echoes influential theoretical work on the nature of *group agency*.

The study of group beliefs, attitudes, and other collective features belongs to the growing philosophical subdiscipline of *social ontology*. While we are accustomed to thinking of individuals as autonomous agents capable of intentional behavior, recent work on the metaphysics of groups suggests that collectives, too, can exhibit varying degrees of agency. One of the first systematic defenses of group agency is due to Margaret Gilbert (1989), whose pioneering analysis has since spawned a rich literature inquiring into the metaphysical basis of group life (Bratman, 2014; List & Pettit, 2011; Miller, 2001; Tollefsen, 2015; Tuomela, 2000, 2013). The field of social ontology now encompasses philosophical analyses of a wide range of agential features, from collective beliefs (Gilbert, 2004; Lackey, 2016) to group motives (Brown, 2021) and shared emotions (Brady, 2016; Huebner, 2011). An important upshot of this literature is that the relation between groups and their members is often complex, such that features instantiated at the collective level can come apart from the features of individual members in philosophically interesting ways.

If groups can exhibit agency in their own right, a question that presents itself to ethical theorists is whether we can appraise the conduct of group agents in the same way that we assess individual conduct. That question has standardly been understood to inquire into the nature of moral responsibility for collective actions, decisions, and beliefs (Bazargan-Forward & Tollefsen, 2020; Feinberg, 1968; French, 1979, 1995; Hindriks, 2009; List & Pettit, 2011;

---

[1] See, e.g., Chozick (2016).

Pettit, 2007). When individuals combine in groups to pursue some collective ends, under what conditions can the group *itself* be held liable for actions performed in name of the group? Take, for example, an organization like the shipping company that operated the ill-fated *Herald of Free Enterprise* (whose deadly capsizing made headlines in 1987). According to official reports, none of the organization's employees bore *individual* responsibility for the ship's sinking. Still, as Philip Pettit (2007) argues, the *organization*'s negligent decisions seem blameworthy. But what does it mean to hold an *organization* responsible for wrongdoing if none of its members are individually at fault?

These are important questions, and the issue of moral responsibility for collective actions, decisions, and beliefs will no doubt continue to generate much scholarly debate. But there is another mode of ethical appraisal that remains relatively unexplored in the extant literature on the ethics of groups. For we can assess moral agents not just on the basis of their individual actions, decisions, and beliefs, but also on the basis of their overall *character* (Watson, 1990). When someone is known for telling pernicious lies, for instance, we do not just judge that person for each individual lie he or she tells, but also for *being a liar*; that is, for the *habit* or *disposition* of lying. This latter mode of ethical appraisal is distinctive in changing the focus of our moral evaluations from the properties of actions to the properties of moral *agents*. Accordingly, it concerns not just the question of what an agent does, decides, or believes, but the question of what *kind* of person or agent they are.

This dissertation takes existing research on the ethics of groups in new directions by critically investigating the conceptual and empirical foundations of an agent-centered approach to group moral theory. Indeed, our ordinary language usage suggests that we frequently evaluate the character of groups in much the same way that we evaluate the character of individual persons. Clinton did not mean to paint the character of *individual* Trump supporters as "deplorable"; she meant to characterize the Trump *movement* as "deplorable." Or consider, again, the shipping company behind the *Herald of Free Enterprise*. We may condemn the company not just for individual acts of negligence, but if these reflect deep-seated habits or patterns of behavior, we may also condemn the company for being a *negligent organization*. Indeed, we judge organizations like this all the time. We blame profiteering corporations for being greedy, and crime syndicates for being callous. On the other hand, we often give groups credit for having a praiseworthy character as well. We praise charities for their generosity, for instance, and commend military squadrons for their bravery.

The main aim of this dissertation is to develop an account of *group character* that is both conceptually rigorous and empirically sound. As we will see, the notion of group

character raises a number of interesting philosophical puzzles. Some of these puzzles mirror those already under investigation in the literature on social ontology. What is the relation between *group* character and the character of individual group members? Under what conditions can groups be held *morally* responsible for instantiating a particular type of character? Other challenges pertain specifically to the idea of group character itself. Since groups are constituted in a radically different way than individuals, showing that groups are capable of realizing equivalent character traits is far from straightforward. In the case of individuals, for instance, the concept of character is usually defined in terms of the *psychology* of individual agents, including a range of distinctive mental states. Are groups capable of realizing these mental states as well? And, if so, how? Does the notion of group character commit us to "metaphysically spooky" (Fricker, 2010, p. 242) claims about "group minds" (Pettit, 2004)? These are just some of the questions we will discuss in the chapters that follow.

In the history of philosophy, the study of character is closely associated with the virtue ethical tradition. That tradition experienced a dramatic upturn after Elizabeth Anscombe's (1958) passionate plea to revive the concept of *virtue* in moral theory. Where moral theorists had long followed the likes of John Stuart Mill, Jeremy Bentham, and Immanuel Kant in trying to articulate the general principles governing our moral universe, Anscombe's seminal paper inspired a new generation of virtue-oriented thinkers to look beyond a principle-based approach to ethics (Foot, 1978, 2001; MacIntyre, 1981; McDowell, 1979; Slote, 2001). The world is too complex, these authors claimed, for a set of moral precepts to tell us what is right and wrong in every circumstance. Accordingly, we should not fixate on the question of whether a certain action accords with moral principles. Instead, these thinkers argued that we should conceive of moral goodness or badness as a function of one's character – in philosophical jargon, of one's *aretaic* qualities. These are qualities that pertain to the praise- or blameworthy character traits and dispositions we call *virtues* or *vices*.

On a standard analysis of virtue, the virtuous agent knows what the right course of action is not because she is competent with a system of rules and principles, but because she has cultivated character traits that reliably lead her to act well (Driver, 2001; Hursthouse, 1999; Swanton, 2003, 2021). Although there is some disagreement as to what these virtuous character traits are, canonical virtues include traits like *courage*, *justice*, and *generosity*. The courageous person, for instance, is someone whose character disposes them to conquer their fears when doing so is required by the circumstances. Conversely, the vicious person is someone whose character is deficient in a way that tends to produce blameworthy actions.

Introduction

Canonical vices include such traits as *cowardice*, *greed*, and *callousness*. Unlike someone with the virtue of courage, a coward *fails* to conquer their fears when doing so is important.

More recently, the concepts of virtue and vice have also found themselves at the forefront of exciting developments in the field of epistemology (Baehr, 2011; Battaly, 2010, 2014, 2016; Cassam, 2016, 2019; Code, 1984; Fricker, 2007; Greco, 2010; Sosa, 1980, 2007; Wood & Roberts, 2007; Zagzebski, 1996). As it turns out, the concepts of virtue and vice lend themselves not just to evaluations of our *moral* character, but also to evaluations of our character as *epistemic* agents. Here, too, is disagreement as to what the (epistemic) virtues and vices in question are. According to one influential conception of epistemic virtue, however, epistemic virtues are character traits that mark an excellent *knower* (Code, 1984; Zagzebski, 1996). Such traits as conscientiousness and open-mindedness, for instance, are epistemically good to have because they facilitate the acquisition of new knowledge. On the other hand, some traits stand in the way of the acquisition of knowledge, and these we can call epistemic vices (Baehr, 2010; Cassam, 2019). A closed-minded person, for example, shuts herself off from acquiring new beliefs and avoids testing the ones she currently holds.

Although the fields of virtue ethics and epistemology have traditionally concerned themselves with virtues and vices as features of individual agents, we have seen that *groups* are frequently evaluated in terms of their character traits as well. This suggests the possibility of applying virtue theoretical insights to group agents. Accordingly, an important objective of this dissertation is to bridge the gap between recent work on group agency and ongoing debates within the field of virtue theory by critically assessing the viability of a *virtue theoretical* approach to evaluating the moral and epistemic conduct of groups. The overarching argument I develop is that group agents can exhibit virtues and vices in much the same way as individual agents. The account of group virtue and vice I develop in this dissertation thus allows us to import novel virtue theoretical insights into the rapidly evolving field of group ethics and epistemology. Just as individual agents can cultivate such virtuous character traits as generosity and justice, I argue, so groups like corporations and political parties can nurture just and generous *corporate* characters. And just as individuals have traits facilitating the acquisition of knowledge, so do corporations: a *curious* organization, for instance, may reliably seek out knowledge and new beliefs by investing in innovative research and development. Perhaps less surprising, we will see that groups are capable of manifesting moral and epistemic *vices* as well.

Crucially, using the language of virtue and vice to appraise group conduct opens up new avenues for dealing with praise- or blameworthy aspects of group life. If the actions of a

certain collective consistently produce morally bad outcomes, a virtue theoretical approach entails that our interest as moral theorists should not just lie with any moral rules that have been violated, but also – and perhaps more importantly – with the deficient character traits and dispositions the collective has acquired over the course of its life. We might remedy that situation by encouraging the collective to develop beneficial habits instead. If, on the other hand, a generally praiseworthy corporate agent produces a one-off bad result, this need not count against the agent's moral character, and accordingly such incidental wrongs should be of less interest to the moral theorist.

Another benefit of adopting the language of virtue and vice is that it is not only evaluative, but also richly *descriptive*. Virtue and vice terms are paradigmatic instances of what Bernard Williams (1985) calls "thick" (p. 130) concepts. While thick concepts evaluate their object *as* good or bad, they also describe *the way* in which their object is good or bad. Calling someone a just person, for instance, evaluates their character as morally praiseworthy, but also describes the specific way in which their character is praiseworthy. Namely, in its disposition to fairness and impartiality. As such, it conveys more information than evaluations using "thin" concepts like *good* or *bad*. Similarly, while calling out a business or organization for being unethical does not convey much information, evaluating a business or organization as *greedy*, *corrupt*, or *irresponsible* describes exactly the respects in which it falls short. Putting this information out there can make it easier to hold group agents accountable for their conduct while simultaneously suggesting concrete areas of improvement.

But I am getting ahead of myself. Before we can reap the benefits of a virtue theoretical approach to the study of group character, I must show that groups are capable of manifesting virtues and vices in the first place. Accordingly, the main research question this dissertation aims to answer is whether and how group agents are capable of manifesting virtuous or vicious collective character traits. I investigate the conditions under which groups of individuals combine to form agents whose character is praise- or blameworthy over the course of five chapters.

My inquiry starts with a detailed analysis of the current state of virtue theory in chapter one. I turn to my analysis of *group* virtue in chapter two. While there is some preliminary work on the prospect of collective virtue and vice (Beggs, 2003; Byerly & Byerly, 2016; Fricker, 2010, 2013, 2020; Jones, 2007; Lahroodi, 2007), my research departs from existing literature in at least two significant ways. For one, this dissertation marks the first systematic attempt to analyze collective virtues and vices as *functional states*. As such, the functionalist account of group character I develop in chapter two contrasts starkly with

existing accounts that largely characterize collective virtues and vices in terms of Gilbertian (Gilbert, 1989) joint commitments (e.g., Byerly & Byerly, 2016; Fricker, 2010, 2013, 2021; Lahroodi, 2007). As we will see, analyzing collective virtues and vices as functional states enables us to study a far wider range of collective virtues and vices than the current literature on group character is able to accommodate. But my research does not just advance the debate on collective virtue and vice *conceptually*. It also departs from the existing literature by studying the *empirical* commitments of theories of group character, both by way of a critical analysis of behavioral economic and social psychological research on group conduct (chapter three) and a detailed case study (chapter four). The result is a novel account of group character that is both conceptually and empirically sound, and provides fertile ground for future research.

I provide a detailed synopsis of each chapter below.

Chapter 1 – Notions of Character: Virtue and Vice
Setting off our inquiry into the aretaic qualities of group life, the first chapter critically examines the current state of virtue theory. Focusing on *moral* theory, it analyzes the roles played by the concepts of virtue and vice in the three dominant approaches to ethics: utilitarian, deontological, and virtue ethical approaches. It also examines two influential conceptions of virtue and vice. On *motivation-based* conceptions of virtue and vice, the moral quality of our character is intrinsically bound up with the quality of our motives (Foot, 1978; Hursthouse, 1999; Zagzebski, 1996). On *ends-based* conceptions of virtue and vice, the moral quality of our character is a function of the ends or consequences that our character helps realize (Driver, 2001, 2016).

Chapter 2 – The Idea of Group Character
Shifting our focus away from *individual* agents and onto *collective* ones, chapter two introduces the idea of *group* character. With ends- and motivation-based conceptions of virtue and vice in hand, it assesses whether and how these conceptions apply to the character of group agents. As a starting point, it considers existing accounts of collective virtue and vice based in Margaret Gilbert's plural subject account of group agency (Fricker, 2010, 2013, 2020; Lahroodi, 2007). Though these accounts offer compelling analyses of *some* collective virtues and vices, I argue that they suffer from two limitations that restrict their applicability to *motivation*-based virtue and vice. This shortcoming is the starting point of my search for an account that accommodates group virtues and vices of all kinds. In the second part of chapter

two, I develop an entirely novel – *functionalist* – account of collective virtue and vice that meets this challenge. On the view I develop, collective virtues and vices arise when corporations or other group actors are organized so to as to *function* as virtuous or vicious agents. This new account of collective virtue and vice can model a broad range of collective character traits, has considerable empirical purchase, and ties in well with empirical literature from the organizational sciences.

Chapter 3 – Expanding the Situationist Challenge to Collectives

Its remarkable rise to philosophical prominence notwithstanding, virtue theory is not without its critics. The most convincing objections to the virtue theoretical project have challenged its empirical commitments. Drawing on empirical work in social psychology, so-called *situationists* argue that virtue theory presupposes character traits that do not in fact exist (Alfano, 2012; Doris, 1998, 2002; Harman, 1999, 2000, 2009). While virtues are character traits that supposedly produce consistent behavior across widely different circumstances, that is, situationists claim that our behavior is in fact highly variable as a result of trivial situational influences. This chapter critically examines the implications of the situationist challenge for my account of group virtue, expanding the situationist challenge in two ways. First, it considers what situationist research tells us about the impact of group dynamics on individual virtues and vices. Second, it investigates whether group *agents* are vulnerable to situationist critiques as well. While I will argue that groups are liable to situational influences much like individuals are, I show that this does not invalidate my account of group virtue.

Chapter 4 – Collective Epistemic Virtue and Vice: The Case of Boeing

This chapter extends a functionalist account of collective virtue and vice to collective *epistemic* virtues and vices. On the view that emerges, epistemically virtuous or vicious groups are organized so as to function *as* epistemically virtuous or vicious agents. The chapter illustrates the practical relevance of this account by way of a case study of the American aircraft manufacturer Boeing. Drawing on media reports and official investigations, it analyzes the *corporate epistemic culture* (de Bruin, 2015) that shaped the conditions leading up to the deadly crashes of two of its Boeing 737 Max jets in 2018 and 2019. Though it would be a mistake to say that these crashes were entirely to blame on epistemic problems, the chapter concludes that collective epistemic vice undoubtedly played a part. The chapter also includes recommendations for fostering an environment conducive to collective epistemic virtue instead.

Introduction

Chapter 5 – Improvised Joint Action

While groups are capable of realizing many of the conditions required for the possession of virtue or vice, the existing literature on group agency does not capture all of these prerequisites. On most conceptions of virtue, for instance, cultivating a virtuous character requires that one is capable of improvising morally good behavior. As I argue in this chapter, however, extant accounts of joint action struggle to explain the possibility of jointly improvised action. The problem lies in the widely held view that jointly intentional behavior requires common knowledge of shared intentionality (Bratman, 2014; Gilbert, 1989; Miller, 2001; Tuomela, 2013). In paradigmatic examples of improvisation, groups are able to improvise joint acts without prior knowledge of shared intentionality. I propose that we can resolve this problem by dropping the condition that common knowledge is required for joint action.

Chapter 1 - Notions of Character: Virtues and Vices

1.1 Introduction

Although the virtue theoretical tradition has a long and varied history dating back several millennia, its modern resurgence is a relatively recent phenomenon. Born from an intense frustration with the doctrines of utilitarianism and deontology, Elizabeth Anscombe's (1958) "Modern Moral Philosophy" is widely credited with restoring the virtue theoretical tradition to prominence.[2] One of the central claims of Anscombe's scathing indictment of the state of moral philosophy is that concepts like *morally ought*, *morally obligated*, and *morally right* are overly legalistic and lack content unless we presuppose the existence of a divine legislator from which these concepts derive their moral authority.[3] Since much of moral philosophy is now secular, Anscombe asks if we can abandon a theistic framework and base our ethical theories on the conceptually rich notion of *virtue* instead.[4] On the kind of moral theory Anscombe envisions, it is not divine law, but the character of moral agents that grounds our ethical judgments. Following Anscombe's example, a new generation of thinkers spearheaded by Philippa Foot (1978), John McDowell (1979), Alasdair MacIntyre (1981), and Michael Slote (1983) has subsequently set the stage for what is the diverse field of virtue theory today.

If Anscombe's problem with modern moral philosophy was the scarcity of work on the nature of virtue, today we are confronted with an embarrassment of riches (see, e.g., Athanassoulis, 2012 for an overview). Surveying the virtue theoretical tradition, MacIntyre (1981) finds so many different and contradictory accounts of the character traits required for a morally good life that he despairingly asks if there is any "single, central, core conception of

---

[2] At least in the field of *analytic* philosophy – see, e.g., Zagzebski (1996, p. 17), Hursthouse (1999, p. 3), and Swanton (2021, p. 3). The field of continental philosophy has a longer history of drawing on the virtue ethical tradition – see, e.g., Jankélévitch (1949) and Pieper (1934). There is also a range of flourishing non-Western virtue traditions, including Buddhism and Confucianism (see, e.g., Vallor, 2016 for discussion).

[3] MacIntyre (1981) makes a similar diagnosis, arguing that *all* our moral vocabulary is incoherent because it presupposes a normative framework now abandoned (see O'Hara, 2018 for comparison).

[4] On an alternative – less prominent – reading of Anscombe's text, her suggestion is not that we embrace virtue theory, but rather that we accept the idea of a divine legislator as foundational and so restore a coherent religion-based moral framework (see, e.g., Blackburn, 2005; Crisp, 2004, p. 86).

the virtues" (p. 173) at all. Urging us to reject the skeptical conclusion that we lack such conception, MacIntyre suggests that we carefully study the dominant accounts of virtue in order to identify the common thread that runs through all of them. That is part of what I aim to do in this chapter. While I take his advice to heart in what follows, however, I depart from MacIntyre in two ways. First, this chapter includes not just a critical analysis of dominant conceptions of virtue, but also of dominant conceptions of *vice*, and so of the character traits that mark a morally deficient agent. Second, I argue that the common themes permeating our conceptions of virtue and vice do not add up towards a single, core conception of the virtues. Rather, I contend they circumscribe a family of different but related conceptions of human excellence and deficiency that we can each fruitfully apply to the character of moral agents.

Setting the stage for my investigation into the aretaic qualities of group life, this chapter critically examines the current state of virtue theory. In section two, I briefly discuss the role the concepts of virtue and vice play in each of the dominant approaches to moral theory: deontology, utilitarianism, and virtue ethics proper. Perhaps surprisingly in view of Anscombe's harsh words, we will see that the revival of interest in virtue theory has seen scholars defend decidedly consequentialist (Coakley, 2017; Driver, 2001, 2016; Hurka, 2001) and deontological (Hill Jr., 2012; Jost & Wuerth, 2011) accounts of virtue and vice. Since virtue ethicists have gone farthest in their exploration of character-based ethics, however, most of the remaining chapter draws explicitly on the virtue ethical tradition. Having clarified the role of virtue in each of the dominant moral traditions, I turn my attention to the *structure* of virtue in section three. Though there is considerable disagreement about the kind of character traits virtues are, I will argue that this disagreement reflects two distinct, but equally important intuitions about moral worth. In section four, I develop a similar argument when I analyze the structure of *vice*. My conclusion sums up and notes the theoretical problems that will occupy us when we try to apply virtue theoretical insights to group agents in the next chapter.

One caveat is in order before I start my investigation into the nature of virtue and vice. Insights from virtue theory have recently been put to use not only in the study of morality, but also within the burgeoning field of virtue epistemology. Indeed, both virtue reliabilists such as John Greco (2010) and Ernest Sosa (1980, 2007), and virtue responsibilists such as Lorraine Code (1984, 1987) and Linda Zagzebski (1996) have made significant theoretical advances by applying the concept of *intellectual* or *epistemic* virtue to the perennial problems of epistemology. More recently, such theorists as Jason Baehr (2010), Heather Battaly (2014, 2016), and Quassim Cassam (2016, 2019) have added to these advances by studying the

hitherto undertheorized concept of intellectual *vice*. While these developments are of tremendous interest, I postpone consideration of intellectual virtues and vices until the chapters that follow. The present chapter focuses on *moral* virtues and vices.[5]

1.2 Virtue Theory

Virtues are standardly understood as morally excellent character traits that *dispose* their possessor to a characteristic type of behavior. The courageous person, for instance, is *disposed* to respond well to fear, danger, or difficulty, taking acceptable risks as required by circumstances. Since the character of a virtuous agent has been of interest to advocates of each of the dominant approaches to moral theory, the study of virtue *theory* will take us well beyond the confines of the virtue ethical *tradition* as understood by Anscombe, MacIntyre, and other early advocates of a virtue ethical approach to moral theory. Anscombe and MacIntyre conceived of virtue ethics as an alternative to the doctrines of utilitarianism and deontology, largely based in neo-Aristotelian thought (see also Foot, 2001; Hursthouse, 1999). But talk about virtue, vice, and the importance of character in moral theory is by no means exclusive to the virtue ethical tradition alone. We therefore need to clarify the role of virtues and vices in each of the dominant moral traditions before starting our analysis of the structure of these character traits in the sections that follow.

The difference between the doctrines of utilitarianism, deontology, and virtue ethics is often characterized in terms of what Battaly (2010) calls the "direction of analysis view" (p. 5).[6] According to this way of characterizing the differences between moral theories, every moral theory relates the concepts of *the good*, *the right*, and *the virtuous* in some particular manner, but theories differ in how they arrange these concepts in order of priority. Utilitarians and deontologists typically derive their account of a virtuous person from their account of right action, which they believe is more basic. For utilitarians and other consequentialists, this means that the virtuous person is someone whose actions consistently produce desirable

---

[5] In what follows, I observe Julia Driver's (2001) distinction between moral, prudential, and intellectual virtues and vices. *Moral* virtues are those that directly concern the well-being of others, *prudential* virtues are those that primarily concern the well-being of the agent who possesses them, and *intellectual* virtues are those that realize the agent's epistemic capacities. There can be significant overlap between these categories (see, e.g., Fricker, 2007), and I will comment on the interesting and complex ways in which the moral, prudential, and intellectual virtues interact where appropriate.
[6] See, e.g., Zagzebski (1996, pp. 78-80), Driver (2001, pp. xiii-xxi), Hursthouse (1999, pp. 1-24), and Swanton (2021, pp. 1-16). For a different characterization of the difference between these approaches, see, e.g., Crisp (2010).

outcomes (Driver, 2001; Hurka, 2001). Deontologists instead maintain that the virtuous agent is someone whose actions reflect the conscientious observance of moral rules (Hill Jr., 2012; Jost & Wuerth, 2011). Virtue ethicists, however, invert this order of importance and derive their account of right action from their conception of the virtuous. Rosalind Hursthouse (1999, p. 28), for instance, argues that an action is right only if it is the kind of action a virtuous person would perform. On this view, the caliber of one's character bestows moral worth upon one's actions, at least to the degree that one's actions are a manifestation of one's character, and not the other way around.

There are broadly two ways in which virtue theorists have related their account of the virtuous to a conception of the good. In what Zagzebski (1996) calls "good-based" (p. 80) theories of virtue, virtues derive their moral value from the connection with some more fundamental good.[7] Both utilitarian and deontological accounts of virtue typically fall into this category. Utilitarian and other consequentialist virtues inherit their value from the quality of their consequences (Coakley, 2017; Driver, 2001, 2016). Accordingly, the utilitarian will count, say, generosity as a virtue only if generosity tends to promote a desirable state of affairs. For classical utilitarians, who conceived of pleasure as the fundamental moral good, that is a state of affairs characterized by the greatest balance of pleasure over pain. The fundamental good from which deontological virtues derive their value, by contrast, is their rootedness in the concern for a moral law (often expressed in a sense of moral duty). For deontologists, then, generosity counts as a virtue only to the extent that it is aligned with the demands that this moral law imposes on moral agents (Hill Jr., 2012; Jost & Wuerth, 2011).

Anscombe's (1958) problem with these good-based accounts is that she believes utilitarians and deontologists are ultimately unable to justify the goods they posit as fundamental. The paradigmatic good-based theories of virtue, however, undoubtedly belong to virtue ethics proper, and face similar problems of justification. In his *Nicomachean Ethics*, for instance, Aristotle (ca. 350 B.C.E./2020) argues that virtues are valuable states of character because they are conducive to, and partly constitutive of, *eudaimonia*, which is usually translated as "human flourishing," "happiness," or "well-being" (Broadie, 2012). According to Aristotle, then, *human flourishing* is the fundamental good from which virtues derive their moral worth. Other philosophers who define the value of virtue partly in terms of human flourishing are Foot (2001), Hursthouse (1999), and Julia Annas (2011). But Aristotle and other *eudaimonist* virtue theorists face some difficulty in explaining just *how* human

---

[7] Good-based theories of virtue are what Slote (2001) calls "agent-prior" (pp. 6-7) theories of virtue.

flourishing confers value on the virtues. If we value virtues because their possession allows flourishing, it seems that we value them only instrumentally, and virtue ethics risks collapsing into a consequentialist moral theory – a result Aristotle and other *eudaimonist* virtue ethicists want to avoid.

Indeed, while Aristotle's account is teleological, or ends-oriented, in that it partly explains the value of the virtues in terms of the ends at which they aim, Aristotle takes great pains to differentiate his moral theory from a purely consequentialist one. Though there is a tension here, Aristotle claims that the virtues are *intrinsically* valuable – that is, independently of their consequences – insofar as they are *themselves* constitutive of the good life, and not merely instrumental in achieving the good life. In other words, human flourishing is not some external good at which the virtues aim; it is a good realized in the exercise of virtuous character traits themselves. Bernard Williams (2002, pp. 90-93) makes a similar point when he argues that such virtues as sincerity and trustworthiness are not just instrumentally useful in facilitating a good life, but intrinsically valuable insofar as valuing these virtues as good in themselves is an indispensable component of a good life.[8] The virtue of trustworthiness, for instance, is not merely instrumentally useful because it enables good forms of communication; it is valuable in an irreducible way given that trustworthiness is the bedrock value without which good forms of communication could never get off the ground. Hence, Williams argues, we *need* to value trustworthiness as an intrinsic good in order to engage in meaningful exchanges of information (ibid., p. 92).

MacIntyre (1981) likewise explores this tension in great detail in *After Virtue*. According to MacIntyre, virtues are instrumentally valuable insofar as they enable us to realize the goods internal to some practice. However, they are intrinsically valuable insofar as they are partly constitutive of that practice and of the goods attained therein. To illustrate, MacIntyre uses the game of chess. In order to reap the rewards internal to a game of chess – the joy of playing, the satisfaction of good strategy, and so on – one must learn to play chess well. But the virtues required for playing chess well are not just instrumentally useful in promoting these goods, they are what *sustain* these goods. An excellent chess player will value the virtue of honesty, for instance, not because an honest game will maximize their chance of winning, but because an honest game is worth pursuing as an end in itself (in other words, is itself an indispensable component of the joy of playing). Since MacIntyre conceives

---

[8] Williams was skeptical about talk of *the* good life. Rather, he believed there are many possible good lives.

of life as a series of practices linked together by various narratives, institutions, and traditions, he concludes on this basis that virtues are valuable because they are indispensable to "realize the good of a whole human life" (p. 189), making him an *eudaimonist* philosopher as well.[9]

One problem with these kinds of good-based accounts is that the meaning of *human flourishing* or *the good life* is notoriously difficult to pin down. Reflecting on what constitutes a good life, Foot (1978) writes, somewhat despairingly:

> [H]ere we meet ideas which are curiously elusive, such as the thought that some pursuits are more worthwhile than others, and some matters trivial and some important in human life. Since it makes good sense to say that most men waste a lot of their lives in ardent pursuit of what is trivial and unimportant it is not possible to explain the important and the trivial in terms of the amount of attention given to different subjects by the average man. But I have never seen, or been able to think out, a true account of this matter. (p. 6)

Much has been written on the good life, yet Foot – well-read, and one of the finest thinkers of the 20th century – has never come across an account that struck her as true. And that is hardly surprising: human beings enact a great variety of forms of life, each of which encompasses different values, norms, and traditions, making it difficult to isolate what *the* good life consists in.[10] This is one of the reasons that MacIntyre, too, struggled to find a single conception of the virtues. Without an account of human flourishing, however, any theory of virtue that defines the value of virtues in terms of the good life remains incomplete, and so without the justification that Anscombe demanded of utilitarianism and deontology.

We can contrast good-based theories of virtue with what Slote (1995, 2001) calls "agent-based" theories of virtue. Avoiding the pitfalls of teleological accounts of virtue, agent-based theories of virtue hold that virtuous states of character are *themselves* fundamental moral goods. On this view, we cannot explain the value of virtuous states of character by appealing to an underlying bottom-level concept such as human flourishing, moral law, or utility. Instead, the agent-based theorist of virtue holds that we must explain the moral value of these other concepts in terms of virtuous states of character. Happiness, the agent-based theorist maintains, is a moral good *because* the virtuous person pursues it, and

---

[9] Like Williams, MacIntyre (1981) believed that "the good of a whole human life" (p. 189) can be realized in many different ways. Though MacIntyre argues that all virtues have a conceptual core in common, he claims that the virtuous lives these virtues sustain do not revolve around a single conception of *eudaimonia*.

[10] This is not to deny, of course, that many have *tried* to give an account of the good life. For an excellent investigations into the capabilities central to living a good life, see, e.g., Nussbaum (2006).

virtues are not moral goods *because* they aim at happiness. Though she does not endorse it outright, Zagzebski (1996, pp. 82-84) expresses considerable sympathy for this view when she weighs the pros and cons of adopting a good- or agent-based theory of virtue. The problem for this kind of theory, of course, is coming up with a non-ad hoc method of showing which of our character traits are valuable in a non-derivative way. Radically, Zagzebski's (2010, 2013) exemplarist theory of virtue contends that we can do this by referring directly to the traits of "moral exemplars," those people throughout history whose traits we find admirable upon reflection. For Slote (2001, p. 38), moral intuitions are our guide to intrinsically valuable character states.[11]

As the above discussion indicates, then, utilitarians, deontologists, and virtue ethicists have each had something to say about the concept of virtue. I have been rather quiet, however, about the concept of *vice*, or about the qualities that characterize a morally deficient agent. As we will see in section four below, this silence reflects the scarcity of work on vice in general.[12] Where theorists do consider vice, they tend to conceive of it simply as the mirror image of virtue (Crerar, 2018). Accordingly, utilitarians hold that vices are traits that are bad because they produce bad outcomes, deontologists hold that vices are bad because they conflict with our sense of duty or with universal moral laws, and virtue ethicists argue that vices are antithetical to human flourishing or bad in themselves. Whether or not this picture is convincing is something I consider later in this chapter.

Before I move on to the next section, I would like to clarify that my aim will not be to vindicate any particular moral tradition. Accordingly, I will try to remain neutral with respect to utilitarian, deontological, and virtue ethical approaches in the hope that the account of group virtue I go on to defend in the next chapter will be of interest to proponents of each of these traditions. Given that virtue ethicists have given the most comprehensive account of the structure of virtue and vice, however, most of the literature I discuss below inevitably draws on the virtue ethical tradition. As Hursthouse (1999) points out, the distinction between the dominant moral traditions is becoming increasingly hard to draw anyway:

> [T]he lines of demarcation between the three approaches have become blurred.
> Describing virtue ethics loosely as an approach which 'emphasizes the virtues' will no
> longer serve to distinguish it. By the same token, of course, deontology and

---

[11] Unsurprisingly, not everyone agrees that intuitions alone are an adequate basis for explaining the value of virtues (see, e.g., Athanassoulis, 2002).

[12] While this is true of moral theory, the concept of vice has recently garnered significant attention within the field of epistemology (see, e.g., Baehr, 2010; Cassam, 2016, 2019).

> utilitarianism are no longer perspicuously defined by describing them as emphasizing
>
> rules or consequences in contrast to character. (p. 4)

Indeed, though the doctrines of utilitarianism, deontology, and virtue ethics attach varying degrees of importance to moral character, they have in common the resources to talk about an agent's character in terms of virtue and vice. And *that*, I submit, is the major appeal of virtue theory. Whether one is a utilitarian, deontologist, or virtue ethicist, the language of virtue and vice enables us to focus our moral evaluations not on the quality of *actions*, but on the character of *agents*. That is, it allows us to judge someone not merely for the individual acts they have performed, but for the *kind of person* they have fashioned themselves into. This is a powerful evaluative resource, particularly because we often tend to believe that someone's character reflects their values, attitudes, and dispositions in a deep and robust manner. To see why this is so – and why not everyone agrees that character traits reflect our agency in this deep way – I now turn to my analysis of the structure of virtue and vice.

1.3 The Structure of Virtue

If the preceding section shows that the concepts of virtue and vice appear in each of the dominant moral traditions, it leaves an important question unanswered. Across these traditions, the power of the concepts of virtue and vice lies in their ability to represent the moral worth of an agent's character in a particular way. But in *what* way? Though we have said that virtues are moral excellences and vices moral failings, we need to say more. Indeed, while some accounts of virtue and vice are more demanding than others, the vast majority of virtue theorists maintain that some of our excellences fall short of virtue and some of our failings are not yet vices. Accordingly, we need to critically examine the different accounts of virtue and vice offered in the literature in order to identify the kind of character traits they are. Only then can we investigate the challenges of applying these concepts to group agents. The present section examines the structure of virtue. The next section focuses on the structure of vice.

Virtues, then, are character traits that make a person morally excellent. But what *kind* of traits are they? We can distinguish between broadly two ways in which virtue theorists have answered this question, each emphasizing a different, but equally valid intuition about moral worth (Battaly, 2010, 2014). The first answer emphasizes the importance of good *motives*. Motivation-based conceptions of virtue hold that the virtuous person is someone

whose character disposes them to do the right thing *for the right reason*, for instance because their values, attitudes, and desires are properly oriented towards the good. This answer captures the intuition that someone's reasons for action often affect our evaluation of them as a person. Morally speaking, for example, it matters whether one is disposed to display kindness to others because one truly cares about benevolence or because one merely wants to keep up the *appearance* of benevolence, perhaps to obtain a certain status within the community.

The second answer emphasizes the importance of good *ends* or *consequences*. Ends-based conceptions of virtue hold simply that the virtuous person is someone whose character is such that they are reliably successful in attaining good ends or producing good consequences. Since ends-based conceptions of virtue do not require that the virtuous person has a virtuous motivation, they are less demanding than their motivation-based counterparts. This answer captures the intuition that, in some domains, bare *performance* is what matters most. According to Henry Sidgwick (1907, p. 201), the justice system is one such domain. In the court of justice, Sidgwick argues, one should not care much about the motives of, say, a prosecutor. Even if a prosecutor is motivated by malice or vindictiveness, what matters is simply that he or she reliably succeeds in upholding the rule of law and that the results of court cases are fair and impartial applications of the law.

It is a matter of some controversy whether motivation- and ends-based conceptions of virtue are mutually exclusive. Their advocates often seem to think so. Foot (1978, 2001), Hursthouse (1999), and Zagzebski (1996), for instance, each claim that motives are an indispensable component of virtues, and so that ends-based conceptions of virtue are mistaken. Driver (2001, 2016), by contrast, is adamant that only ends-based conceptions of virtue can ultimately do justice to our moral intuitions. My contention, however, is that we need not choose one conception of virtue over the other. Rather, we can follow Battaly (2015) in embracing *pluralism* about virtue. As Christine Swanton (2003, 2021) also emphasizes, virtuous conduct is characterized by a variety of different modes of responding to the demands of morality. Accordingly, there is a variety of different ways in which one may achieve moral excellence. Some of these will involve motives, as when an attitude of love moves us to benevolence, while others will privilege outcomes, as when a prosecutor is instrumental in achieving justice despite contravening motives. It stands to reason, then, that our moral excellences do not correspond to a "single, central, core" (MacIntyre, 1981, p. 173) conception of the virtues, but that different conceptions of virtue pick out different types of

moral excellence. We can switch between these conceptions as appropriate. I will now discuss the details of motivation- and ends-based conceptions of virtue in turn.

1.3.1 Motivation-based Conceptions of Virtue

On motivation-based conceptions of virtue, a virtuous agent is someone whose character disposes them to act from virtuous motives (Foot, 2001; Hursthouse, 1999; Zagzebski, 1996). This is usually taken to entail that the virtuous agent is someone who reliably does the right thing *for the right reason*, for instance because they have cultivated the right values, attitudes, and desires – in short, a proper orientation towards the good. Someone with the virtue of honesty, on this conception, is someone who is *disposed* to honest behavior for the reasons that make honesty a morally excellent trait. What these reasons are will depend in part on ones prior (meta)ethical commitments, but on one plausible construal of honesty it requires simply that one values honesty (e.g., telling the truth, not distorting the facts, etc.) *for its own sake* (Miller, 2020). The picture that emerges, then, is of a person who is disposed to perform honest actions simply because it is the virtuous – *honest* – thing to do.

Advocates of motivation-based conceptions of virtue often give these character dispositions what amounts to a standard counterfactual analysis. On this analysis, if someone is disposed to a certain type of behavior, *p,* this means that they *would* display *p* if they were in appropriate circumstances. For someone to possess the virtue of honesty, for example, it is not sufficient that they act from honest motives on one-off occasions. Rather, they must be disposed to act from honest motives across the range of situations in which honesty is an important value. In the language of social psychology, the virtuous person must display "cross-situational consistency" (p. 127) in their behavior (Ross & Nisbett, 2011). The honest person does not just play fair in games of chess, she is *consistent* in her honesty, and so also displays honesty in other types of situations. If the opportunity arises to cheat on a math test, the honest person would therefore refuse to take it. Virtues, in short, are *robust* dispositions to act from virtuous motives.

As we will see in chapter three, conceiving of virtues as cross-situationally robust traits of character has exposed virtue theorists to fierce empirical objections. Indeed, so-called *situationists* charge that social psychological evidence suggests cross-situationally robust character traits do not exist (Doris, 1998, 2002; Harman, 1999, 2000). While I will argue that these objections can be met, for present purposes this dispositional analysis of virtuous

character traits helps highlight another contentious aspect of motivation-based conceptions of virtue. That is the distinction between acts *of* (or *from*) virtue and acts that are merely *in accordance with* virtue.

Acts that are merely in accordance with virtue are consistent with the demands of virtue, but do not express a virtuous motivation. The account Hursthouse (1999) gives of right action, for instance, leaves open the possibility that non-virtuous persons can act in a morally correct way. According to Hursthouse, an action is right if and only if "it is what a virtuous agent would characteristically (i.e., acting in character) do in the circumstances" (p. 29). Suppose, then, that someone who is ordinarily disposed to spread vicious lies occasionally breaks character and acts exactly as an honest person would. That is, she occasionally tells the truth simply because it is the right thing to do. Then even though that person occasionally performs right actions (and is to that extent praiseworthy), this is not yet indicative of a virtuous motivation. Indeed, the concept of motivation employed by advocates of motivation-based conception of virtue is significantly more demanding.

Acts *of* (or *from*) virtue, by contrast, *do* express a virtuous motivation (Zagzebski, 1996, p. 248). As such, they arise directly from someone's virtuous states of character and are more deeply praiseworthy than acts that are merely in accordance with virtue. Having a virtuous motivation is not just a matter of acting for the right reasons *sometimes*, it is what Zagzebski (1996) describes as "a persistent tendency" (p. 132) to act for virtuous motives. The virtuous person's motives, writes Hursthouse (1999), give reasons for action that "go all the way down" (p. 123). These motives are the manifestation of a fundamental ethical outlook that explains the virtuous agent's conduct not just on one-off occasions, but across large chunks of her life. When we ascribe a virtuous motivation to someone, then, we ascribe something "that goes far beyond the moment of action" (ibid.). If someone with a virtuous motivation tells the truth, it is not a fluke; it is what Zagzebski (1996) calls a "deep and important" (p. 85) part of the virtuous person's identity. In other words, the source of the virtuous person's motives is in her deeply entrenched values and dispositions; that is, in those "settled state[s] of good character" (Hursthouse, 1999, p. 123) that we call "the virtues."

The distinction between acts of virtue and acts that are merely in accordance with virtue has proved contentious for broadly two reasons. First, it appears that this distinction does not always track morally salient features of the action under consideration. Sidgwick (1907), recall, argues that there is a class of actions whose moral worth does not depend on the agent's motivation whatsoever. Driver (2001, 2016) makes a similar argument. In line with the virtue pluralism we discussed earlier, this seems to speak in favor of embracing a

plural conception of virtuous action: *some* acts of virtue require a virtuous motivation, *others* do not. Second, philosophers have objected that this account of virtuous action is too demanding (Broadie, 1991, p. 198; Driver, 2001, 2016; see Hursthouse, 1999, pp. 136-140 for discussion). If virtuous action requires that we act from virtuous motives that are deeply entrenched in our settled states of character, virtuous action may not be something that many people are capable of. After all, how many people reliably perform acts of courage, honesty, and so on because they have genuinely internalized these values as constitutive elements of their identity? Further, this conception of virtuous action might seem to block off the possibility for redemption. If virtuous action requires having a virtuous character, then how can the non-virtuous person hope to become virtuous?

While we will revisit this demandingness objection later, we can defuse it in part by studying the way in which advocates of motivation-based conceptions of virtue argue that a virtuous character is acquired. In order to do this, it is instructive to compare virtues with excellences that fall short of full moral virtue. Consider, for example, natural faculties (e.g., good eyesight), innate talents (e.g., having a good ear for music), capacities (e.g., intelligence), and temperament (e.g., a kindly disposition). These are traits that we often admire in people, and they are frequently instrumental in producing morally good outcomes. Someone with good eyesight, for instance, may spot the beginnings of a wildfire long before it has a chance to spread, potentially preventing much suffering. And for someone with a kind temperament, spreading kindness to others without a hint of ulterior motives may come naturally even when they are a young child. The vast majority of virtue theorists, however, follow Aristotle in maintaining that such natural traits cannot be fully-fledged virtues.

One reason virtue theorists have withheld the status of virtue from natural traits is that virtues are often thought to be *praiseworthy* states of character. Virtues, Battaly (2014) writes, are a "credit to us" (p. 54). When we ascribe the virtue of, say, benevolence to someone, we appear to praise that person for the kind of character they have. This implies that virtues are traits for which we hold someone morally responsible. The fact that we hold someone responsible for their virtues entails that they are, in a sense to be specified, *voluntary* traits. Indeed, barring special cases, we neither praise nor hold responsible someone for things wholly outside their control. But whether we are born with good eyesight, innate musical talent, intelligence or a kindly disposition is entirely a matter of what Williams (1981) calls "constitutive luck" (p. 20). That is, natural traits are fully the product of influences outside our control, such as the genetic predispositions of our parents. Accordingly, we can no more praise someone for the presence of these traits than we can blame them for their absence.

What, then, is the sense in which virtues are voluntary traits? That question can be difficult to answer, particularly in view of Oscar Wilde's truism that "at fifty, everyone has the face he deserves." Most advocates of a motivation-based conception of virtue argue that virtues are acquired traits, but claim that a virtuous motivation cannot be acquired with "the flip of a switch" (Zagzebski, 1996, p. 196). Rather, a virtuous motivation must be cultivated over time. While natural traits are the "raw materials" (p. 104) from which we can construct a virtuous personality, Zagzebski (1996) writes, we must work to transform them into full moral virtues.[13] Indeed, the virtuous person has worked to entrench the natural traits that dispose them to virtuous conduct and resist those that run contrary to virtue, affirming and denying these traits in the choices made over a lifetime. Thus, though somebody cannot be held responsible for their natural traits, talents, and so on at birth, they *can* be held responsible for developing or neglecting these traits afterwards (see also Battaly, 2019).

According to Zagzebski, virtues are deeply praiseworthy precisely because they are genuine achievements that require hard work – a sustained history of making the right choices – to develop. This position may initially seem to be at odds with Foot's (1978) stance on natural traits. For Foot, natural traits fall short of full moral virtue not because they are "too easily acquired," but because natural traits "could be disastrous without justice and wisdom, and *these* virtues have to be learned" (p. 11, emphasis mine).[14] The reason natural traits cannot be full moral virtues according to Foot, then, is that they are apt to be harmful if we do not learn how to control them by cultivating the virtues of justice and practical wisdom. Foot here echoes David Hume's distinction between *natural* and *artificial* virtues (Hume, 1779; Wiggins, 1998). Natural virtues, Hume argues, are those traits and sentiments that naturally dispose us towards morally praiseworthy behavior, such as the virtue of having a benevolent spirit. However, these virtues do not necessarily serve us well when it comes to the artificial conventions that govern modern societies. That is why natural virtues must mix with *artificial ones*, which must be learned and dispose us towards morally praiseworthy behavior in the context of societal conventions. Justice, Hume claims, is one such virtue, because it presupposes such conventionally agreed upon norms as e.g. the right to property.

---

[13] *Cf.* Annas (2003): "Virtue is a skill, in the virtue tradition, the skill of living your life in a way which turns your raw materials into a life lived with and from understanding" (p. 27).

[14] In the background of Foot's assertion is a version of the now widely rejected thesis that there must be unity among the virtues (cf. Toner, 2014), such that a trait only counts as a virtue if it harmonizes with all other virtues. Most virtue theorists now hold that one can possess a virtue without possessing all other virtues.

Given the significant work that goes into developing the virtues of justice and wisdom, however, Foot's position is closer to Zagzebski's than it appears. In order to see this, we must briefly turn to Aristotelian doctrine. Foot's claim that the possession of full moral virtue requires the virtue of practical wisdom (*phronesis*) hearkens back to Aristotle's (ca. 350 B.C.E./2020, book VI) denial of the possibility of virtue in children. Aristotle's argument unfolds roughly as follows. Just as "a strong body which moves without sight may stumble badly because of its lack of sight" (ibid., 1144b), so someone without practical wisdom may fail to exercise their natural virtues in accordance with the demands of morality. Suppose, for example, that someone is innately disposed towards *courage*, such that they are not easily afraid. That person may well put their courage to all sorts of vicious ends, as when they display courage in the commission of, say, a daring heist. But in the person whose natural courage enables such vicious crimes, courage is not a virtue (Foot, 1978). What that person lacks is a reliable connection between her courageous disposition and doing the right thing. This suggests that part of having a virtue is being sensitive to the demands of that virtue on different occasions, including the ability to discriminate between ends worth pursuing and ends that are not. Acquiring that sensitivity – for Aristotle, a crucial component of practical wisdom – takes time and experience. Since children come into this world *without* that experience, they cannot be virtuous.

Most advocates of motivation-based conceptions of virtue agree with Aristotle that we acquire a sensitivity to the demands of virtue in roughly the same way that we learn a skill, namely, through habituation (Hursthouse, 1999, 2007; MacIntyre, 1981; Zagzebski, 1996). To return to MacIntyre's (1981) chess example, one learns to become a skilled player of chess through *practice*. As a novice, this involves studying the rules of chess by rote, following the guidance of role models, mechanically applying their advice, and learning from the many mistakes one inevitably makes. With practice, one can gradually let go of these crutches. As one gains a deeper understanding of the dynamics involved, one can stop trying to copy good plays by others and imaginatively devise a strategy of one's own. In time, one will also gain an appreciation of the ends and values that make chess worth playing and start desiring these for their own sake: its competitive intensity, the satisfaction of good strategy, and so on.

So, too, with virtue. In order to cultivate the virtue of courage, we must practice being courageous. Initially, this may involve studying a set of moral rules ("conquer your fears!") and copying the behavior of courageous role models in our community. Since we will lack a comprehensive understanding of the courageous person's reasons for acting, we will frequently make mistakes and use these as input for learning. In so doing, we will become

increasingly attuned to the demands of courage across different circumstances, gradually acquiring the values, attitudes, and desires appropriate to a courageous person. We will learn, for example, to distinguish courage from rashness, and to identify which opportunities call for courage and which require caution. Eventually, acting courageously becomes a habit and our disposition to behave courageously will be entrenched, a "settled state of good character" (Hursthouse, 1999, p. 123). Once our mastery of courage is fully autonomous – i.e., when we have learned how the demands of courage fit together, when this has become part of our ethical outlook, and when we reliably act accordingly – we will have fully acquired the virtue of courage.

While we acquire virtues in much the same way as skills, however, advocates of motivation-based conceptions of virtue argue that virtues are different from skills in important ways. The central difference to which they draw our attention is that we can motivationally detach ourselves from skills in a way that is impossible with virtues. Though skills *are* excellences, their value is purely instrumental. Skills, that is, may help us realize valuable as well as disvaluable ends. I can be skilled at tying my shoelaces or shoplifting, though there is little moral value to be had in doing so. By contrast, virtues are intrinsically bound up with the moral goods at which they aim. Once we recognize the importance of these goods, they make an unconditional claim on our motivation. Broadie (1991) illustrates this by observing that while there is no contradiction in choosing to repudiate a skill, the same is not true of virtues. Thus, there is no contradiction in, say, a skilled piano player choosing to abandon the life of a musician if it no longer interests them. There *is*, however, a contradiction in a benevolent person choosing to repudiate the value of kindness, for by definition a benevolent person would recognize the value of kindness as too important to give up.

Let us finally return to the objection that motivation-based conceptions of virtue are too demanding. On motivation-based conceptions of virtue, we have seen, virtues are acquired traits of character that require significant work to develop. As such, cultivating a virtuous motivation is a tall order indeed. But advocates of motivation-based conceptions of virtue are adamant that virtue is accessible to anyone willing to expend the time and effort trying to achieve it. In the words of Foot (1978), "there belongs to [moral] wisdom only that part of knowledge which is within the reach of any ordinary human adult human being" (p. 7).[15] The motivation of the virtuous person is *not* the motivation of saints; it is the motivation

---

[15] We must be careful not to construe the term "ordinary adult human being" too narrowly. Any plausible account of virtue must allow that virtue is accessible to persons with e.g. disabilities (Nussbaum, 2006).

anyone can acquire over the course of an ordinary human life; and even vicious persons can attempt to redeem themselves by striving to cultivate virtuous dispositions instead.

Moreover, even when someone falls *short* of full moral virtue in this motivationally demanding way, that person can be praiseworthy for *approximating* a virtuous motivation. Virtues, on this motivation-based conception, can come in degrees (Hursthouse, 1999, p. 143). While full moral virtue requires that one's virtuous dispositions are firmly entrenched, someone who is disposed to act virtuously *most of the time* is closer to full moral virtue than someone who has failed to cultivate virtuous dispositions at all. Accordingly, while they are not praiseworthy in the same sense as a fully virtuous person, their character merits more praise than the person who lacks virtuous dispositions altogether. Though motivation-based conceptions of virtue are demanding, then, they hardly represent impossible ideals, and they hardly leave us without the resources to praise good behavior that falls short of full moral virtue.

In short, this motivation-based conception of virtue holds that virtues are acquired character traits that robustly dispose someone to act from virtuous motives. They are deeply praiseworthy given the significant work that goes into cultivating the wisdom, values, attitudes, and desires that entrench these dispositions firmly in their possessor. As we will see now, *ends*-based conceptions of virtue lack this motivational component and are therefore significantly less demanding.

1.3.2 Ends-based Conceptions of Virtue

On ends-based conceptions of virtue, virtues are character traits that reliably facilitate success in attaining morally good ends or producing morally good consequences. Since this conception allows that an agent may be virtuous regardless of their motives, it is less demanding than its motivation-based counterpart. Suppose, for instance, that someone reliably donates to charity for purely egoistic reasons – say, to satisfy a narcissistic desire to be perceived *as* generous. While that person's non-virtuous motives would count against their possession of the virtue of generosity on a motivation-based account, ends-based conceptions of virtue have no problem ascribing the virtue of generosity in this scenario. After all, if the charity in question is successful in, say, alleviating suffering or educating disadvantaged children, the person in question reliably succeeds in promoting the *ends* of generosity.

Whereas motivation-based conceptions of virtue find their inspiration in (neo-) Aristotelian thought, ends-based conceptions of virtue trace their roots to Platonic (Battaly, 2010, p. 3) and Stoic (Annas, 2010) conceptions of virtue. The most influential contemporary proponent of an ends-based conception of virtue is Driver (2001, 2016). On the "minimal" account of virtue Driver (2016) defends, virtues are any and all "character traits that systematically produce more actual good than not" (Driver, 2001, p. 68). According to Driver, then, having such internal states as morally praiseworthy motives is neither necessary nor sufficient for the acquisition of virtue.[16] Rather, what matters is that one's character traits systematically lead to more morally good consequences than bad ones. For Driver, in short, generosity is a virtue only if the objective state of the world is better off *with* generous persons than without. In this, she is an *objective consequentialist* (Railton, 1984). To determine whether a trait is virtuous, Driver claims, it is irrelevant whether its possessor *believes* it brings about good consequences; only the consequences that actually manifest themselves determine its status as a virtue.

The central intuition that ends-based conceptions of virtue accommodate is that our bare performance with respect to morally desirable ends often carries significant moral weight. Consider, again, the generous person: even if she does not donate to charity out of the goodness of her heart, it is conceivable that many people are helped through her donations. Isn't that, Driver asks, morally significant in its own right? This emphasis on consequences, however, also invites a range of important objections.

The most pressing objection consequentialist accounts of virtue must overcome involves the problem of moral luck (Nagel, 1993; Williams, 1981). Since influences outside our control can influence the consequences of our character traits, whether or not we count as virtuous on this conception is partially a matter of chance. Suppose, for example, that a generous person donates her money to a legitimate charity that, unbeknownst to her, turns out to be involved in various immoral activities. Such mishaps do, occasionally, happen: in recent years, the charity Oxfam has been implicated in various sexual misconduct scandals, with staff members allegedly using Oxfam properties to solicit the services of (possibly underage) sex workers.[17] In scenarios such as these, a person's generosity may contribute to morally bad consequences she was not aware of and had no control over. If it turns out these bad

---

[16] This does not entail that motives are necessarily *irrelevant* to having a certain virtue. If having a certain motivation is instrumental in producing morally good outcomes, that motivation would hold moral value on this account too.

[17] For details on the Oxfam allegations, see BBC News (2018, 2021) and Oxfam International's (n.d.) response.

consequences outweigh the good ones, a consequentialist would conclude that that person's generosity is not a virtue but instead a vice. The fact that the designation of our character traits as virtuous or vicious is hostage to circumstances outside of our control, however, flies in the face of the intuition that they are traits for which we hold someone *morally responsible*.

Driver (2001) has sought to minimize the impact of moral luck on her account of virtue by including the qualification that virtues are traits that "systematically" (p. 61) lead to morally beneficial outcomes. This means that we must evaluate the consequences of a certain character trait not on an individual basis, but across the population as a whole (ibid., pp. 82-3). To return to our example, if it turns out that *most* generous persons who donate to charity reliably produce more good consequences than bad ones, such that the overall good produced by generosity outweighs the overall bad, generosity is a virtue even in the person whose generosity incidentally happens to lead to morally bad consequences. While this reduces Driver's exposure to the problem of moral luck, however, she ultimately has to bite the bullet. If a trait systematically leads to morally bad consequences outside of our control, then on Driver's account it is not a virtue. Thus, Driver claims, "the tendency to be blindly charitable, that is, to see the good in others and not the bad," may not have been a virtue in "Nazi Germany," where complacency was apt to perpetuate grave injustices, though it arguably was one in "19th-century rural England" (p. xvii), where charitability may have contributed to harmonious village life, even if it is entirely a matter of luck which society we are born into.

Minimizing the role of moral luck is one of the reasons motivation-based conceptions of virtue attach such importance to the *motives* of a virtuous person, the thought being that our motives are something we *do* have significant control over. Though this move is largely successful, it should be noted that motivation-based accounts of virtue are not totally immune from the problem of moral luck either. For one, the motives we cultivate and values we acquire are often partially a function of our environment (Battaly, 2016; Begby, 2013; Fricker, 2007). For another, most motivation-based conceptions of virtue include a success component too. Indeed, Zagzebski (1996) explicitly defines virtue as a "success term" (p. 136). Foot (1978) likewise observes that "failures in performance rather than intention may show a lack of virtue" (p. 4). The underlying idea here is that insofar as the virtuous person is someone who is motivated to realize good ends, *achieving* these ends is part of virtue. Someone whose motivation is sound, then, but who consistently fails to realize good ends is someone who fails to achieve important moral goods, even if the failure in performance is due to bad luck. Since motivation-based conceptions of virtue attach greater importance to the

quality of our motives, however, it is less vulnerable to moral luck than its ends-based counterparts.

Against motivation-based conceptions of virtue, philosophers like Driver commonly object that they are too demanding. We have already seen a version of this objection when we discussed the significant work that goes into acquiring a virtuous motivation. A second version of the demandingness objection is that motivation-based conceptions of virtue are *intellectualist*. According to Driver, motivation-based conceptions of virtue put too high a premium on the role of wisdom and moral knowledge. After all, most motivation-based conceptions of virtue hold that acting from virtuous motives requires that one *knows* how to identify and respond wisely to the morally salient features of any given situation. This seems to imply that the virtuous agent must *know* that she is doing the right thing.

However, Driver (2001) claims, there is a range of virtues – the "virtues of ignorance" (p. 12) – for which that knowledge is either unnecessary or downright problematic. The virtue of modesty, for example, seems to involve *mis*representing (i.e., deflating) the value of one's achievements, and it is often said that true modesty requires not knowing that one is modest (Driver, 1999). Or take the virtue of blind courage. Suppose someone impulsively runs into a burning building to save those trapped inside. To the extent that this person "fails to perceive any danger to himself, he isn't overcoming any fear or sense of danger" (Driver, 2001, p. 33), essentially failing to register a morally salient feature of the situation at hand. If, as advocates of a motivation-based conception of courage hold, courage is a *motivated* defiance of danger that involves a certain degree of moral awareness, writes Driver, this person does not count as courageous. But, she argues, that seems wrong: regardless of that person's motives, saving persons from burning buildings seems like a morally good thing to do. According to Driver, it is therefore a mistake to insist that virtues necessarily involve such intellectual excellences as wisdom and moral knowledge.

The objection that motivation-based conceptions of virtue are intellectualist, however, is not very convincing for all virtues. Even if we agree that the virtue of modesty requires a degree of ignorance, there are good reasons for doubting that blind courage should count as a virtue. Motivation-based conceptions of courage hold that the courageous person has cultivated a responsible orientation towards danger and knows how to distinguish courage from the vices of cowardice and recklessness. In some cases, then, doing the courageous thing requires *not* taking a particular risk, as when one lacks the training or equipment necessary to have any reasonable hope for a good outcome. A person who consistently acts on impulse and fails to perceive danger is arguably unable to reliably distinguish situations that are too risky

from those that are not, and so lacks the virtue of courage. This seems right. Someone who runs headlong into a burning building without heed of the dangers involved is more likely to suffer the vice of *recklessness* instead, revealing an improper orientation towards risk, with potentially devastating consequences.

For this reason, I wonder if impulsive courage should count as a virtue even by Driver's own lights. Does a trait like impulsive courage really "systematically produce more good than bad"? That is an imponderable question, but it is far from obvious that the balance of consequences weighs in favor of blind courage. Surely if we all went off running into burning houses without due regard for the dangers involved, much tragedy would result alongside the lives that would be saved. Though the question of whether the good consequences outweigh the bad ones is an empirical question we may never be able to answer, it seems plausible that many an ostentatiously courageous person would suffer injury to life and limb if they failed to perceive danger altogether. Hence, Driver's conclusion that courage counts as a virtue even when it is not informed by any perceptions of, or beliefs about, risk and danger appears premature.

Further, Driver's charge of intellectualism is based on a rather narrow interpretation of "virtuous motives." In her discussion of impulsive courage, Driver suggests that someone counts as virtuous in a motivationally demanding way only if her actions are the product of conscious deliberation. If this is correct, Driver's impulsively courageous person does not count as virtuous because the danger to themselves "doesn't figure into [their] deliberations at all" – they "may just act […] without pausing to register salient facts of the situation and weigh alternatives" (Driver, 2001, p. 34). Most sophisticated accounts of motivation-based conceptions of virtue, however, can accommodate cases in which an agent does the right thing without conscious deliberation. Neither Foot, Hursthouse, nor Zagzebski, for instance, believes that virtuous agents should "pause" to "deliberate" about the best course of action in every circumstance. These philosophers hold that it is precisely the mark of a virtuous person that they can often dispense with explicit practical deliberation; their virtuous dispositions are so entrenched that they are like second nature (recall Hursthouse's [1999] words that having a virtuous motivation goes "far beyond the moment of action" [p. 123]). Indeed, it is plausible that the truly courageous person can size up the danger represented by a particular situation in the blink of an eye, and she can do this because her prior experiences have taught her a proper orientation towards risk. It is for this reason that McDowell (1979) refers to virtue as "a sort of perceptual capacity" (p. 332). Through practice, experience, and habituation, the virtuous person often just *sees* what the right course of action is *without* consciously weighing

28

alternatives, and *that* is her reason for action. Contra Driver, then, motivation-based conceptions of virtue do not entail that a virtuous person must know that they are doing the right (that is, virtuous) thing at the level of conscious deliberation.

While Driver's objections to motivation-based conceptions of virtue are ultimately unsuccessful, the general point that in judging someone's moral character we often attach significant weight to the consequences of their actions is not without merit. That is, while motivation-based conceptions of virtue are neither too demanding nor necessarily intellectualist, there are spheres of action in which we care more about bare performance than underlying motives. If Sidgwick (1907) is correct, *justice* is one such sphere; perhaps generosity is another. In view of this, it seems plausible that ends- and motivation-based conceptions of virtue are not mutually exclusive, giving incompatible accounts of the *same* virtues, but complementary, giving accounts of *different* virtues (Battaly, 2015). Since there are different ways in which we can realize moral excellence, ends- and motivation-based conceptions may *both* be part of a family of different but related concepts of moral excellence. Our task will be to find out which virtues are more plausibly construed along ends-based lines and which along motivation-based ones.

Let us briefly take stock. We have seen that the concept of character plays a role in each of the dominant moral traditions and critically examined two prominent conceptions of virtue. On motivation-based conceptions of virtue, the virtuous person is someone who is disposed to act from virtuous *motives*. On ends-based conceptions of virtue, the virtuous person is someone who is disposed to realize virtuous *ends*. Since both conceptions of virtue reflect important moral intuitions, I will not seek to vindicate one conception of virtue over the other, and so proceed on the assumption of pluralism in what follows. The differences between these conceptions of virtue will become important, however, when I investigate the concept of *collective* virtue down the road. Indeed, one of the main objectives of this chapter is to analyze the structure of virtue such that I can assess the challenges in showing that virtues can be realized by collectives. Given the fact that motivation-based conceptions of virtue impose more stringent demands on the psychology of the virtuous agent, it will be more difficult to show that groups are capable of realizing motivation-based virtues than ends-based ones. But I am once again getting ahead of myself, for in order to develop a comprehensive account of group character we must also study the structure of *vice*. It is to that task I turn next.

1.4 The Structure of Vice

Since this dissertation also aims to offer an account of collective *vice*, we must briefly consider the structure of vicious character traits as well. Although Anscombe's bid to restore the concept of virtue to philosophical prominence has undeniably been successful, the concept of moral vice remains vastly undertheorized. Where virtue theorists *do* address vice, they often treat it as an afterthought. If virtues are character traits that make us excellent, then vices are those that render us deficient – and not much more is said. This discrepancy is odd, particularly in view of the fact that we appear to possess many more concepts for *vices* than we do for virtues. In the words of Hursthouse (1999), "it is a noteworthy feature of our virtue and vice vocabulary that although our list of generally recognized virtue terms is […] quite short, our list of vice terms is remarkably – and usefully – long" (pp. 41-2).[18] While Hursthouse's empirical claim about the content of our moral vocabulary is worthy of further investigation (in e.g. the field of linguistics), the abundance of vice terms raises an important question. Why has the concept of moral vice escaped the rigorous scrutiny reserved for the concept of virtue?

Though it is difficult to pinpoint the source of this discrepancy, one sometimes gets a sense that it is due to a general exhaustion with the emotion of blame. We tend to think of vices as *blameworthy* character traits, and Linda Zagzebski (1996), for instance, believes that "the practice of blaming is vastly overdone," expressing a deep sympathy for the "desire to tame the practice of blaming" (p. 258). She thereby joins H. H. Price (1954) (whom Zagzebski cites approvingly), who complains that "we are all far too much addicted to blaming people," warning for a "perfect orgy of moral indignation and condemnation" if we were to blame individuals even "for the way they direct their thoughts" (p. 23), i.e., for their psychologies. In its focus on moral failings, the study of vice is entirely *negative*. Zagzebski and Price choose instead to focus on the *positive*, including, in Zagzebski's case, on the moral and intellectual excellences we should aspire to.

Tired as we may be of the practice of blaming, however, the social problems in which vice plays a role give us a strong reason for wanting to deepen our understanding of the

---

[18] Hursthouse (1999) goes on to list our vice terms: "[m]uch invaluable action guidance comes from avoiding course of action that are irresponsible, feckless, lazy, inconsiderate, uncooperative, harsh, intolerant, indiscreet, incautious, unenterprising, pusillanimous, feeble, hypocritical, self-indulgent, materialistic, grasping, short-sighting, … and on and on" (pp. 42-3). Meanwhile, discussions of virtue typically focus on a narrow set of cardinal virtues including justice, courage, and generosity.

concept of vice.[19] While the project of virtue theory has been to present us with ideal states of character, many of us fail to realize these character states in our lifetime. Though we will see below that not every character trait that falls short of virtue constitutes a vice, much havoc is wrought by such vicious character traits as negligence, greed, and cowardice. As instructive as it is to know what ideals we should aim for, it is therefore equally important that we get a grasp on the many ways in which we can fall short of these ideals. That is one of the insights underlying Fricker's (2015) "failure-first methodology" (p. 73). As Fricker points out, *understanding* vices and other moral failings is often the first step towards mitigating them. Thus, even *if* we are primarily interested in virtuous conduct, the study of vice is important because it will reveal which vicious pressures we ought to stave off.

Another reason virtue theorists have not given as much attention to the concept of vice may be the influence of what Charlie Crerar (2018) calls "the inversion thesis" (p. 764). The inversion thesis is the widespread assumption among virtue theorists that vice and virtue are symmetric character traits, such that what holds true of virtue is true of vice in the opposite direction. If vice and virtue mirror each other exactly, Crerar writes, then "the lack of attention to vice would be justifiable," because "a theory of vice would fall neatly out of a theory of virtue" (ibid.). All we would need to do to construct a theory of vice is invert the logic underpinning our theories of virtue. That is what Battaly (2014) does after suggesting that we turn to the dominant conceptions of virtue and "use their underlying rationales to develop corresponding concepts of vice" (p. 56). But at least in the domain of virtue *epistemology*, Crerar argues, there are interesting asymmetries between the concepts of virtue and vice. In view of these asymmetries, Crerar concludes that the inversion thesis is wrong, and so that the lack of attention to vice is *un*justifiable.

While Crerar focuses on *intellectual* vices, his warning has important implications for moral theorists as well.[20] For if an analysis of moral vice shows that it is *not* strictly the mirror image of moral virtue, then the dearth of work on moral vice is a genuine theoretical shortcoming. In what follows, I argue that there are in fact some asymmetries between the concepts of moral vice and virtue, though these do not stop us from constructing conceptions of vice that roughly mirror the conceptions of virtue discussed earlier. Accordingly, I first

---

[19] There are equally strong reasons to reject the notion that the practice of blaming is "vastly overdone." In many cases, blaming individuals is a first step towards acknowledging and remedying moral transgressions. See, e.g., Fricker's (2016) argument that the practice of blaming can be constructive in communicating and negotiating moral norms and values.

[20] Despite his focus on intellectual virtue, Crerar (2018) suspects that "similar conclusions […] will be applicable to virtue *ethics*" (p. 755, emphasis mine).

discuss *motivation-based* conceptions of vice and then turn to *ends-based* conceptions of vice in what follows.

1.4.1 Motivation-based Conceptions of Vice

There is a straightforward sense in which vices *are* the exact opposites of virtues. While virtues are moral excellences, vices are moral deficiencies. On motivation-based conceptions of vice, this deficiency is primarily a deficiency in motivation. When Foot (1978), Hursthouse (1999), and Zagzebski (1996) write about vice, they largely develop this conception of vice analogously to their conception of virtue. On the picture that emerges, vices are motivated character defects that, in Hursthouse's (1999) words, "go all the way down" (p. 123). Whereas the virtuous person has cultivated a fundamentally *good* moral outlook that robustly disposes her to act well (and from virtuous motives), the vicious person has cultivated a fundamentally *corrupt* moral outlook that robustly disposes her to act badly (and from vicious motives). As in the virtuous person, the vicious person's moral outlook is so entrenched that it reliably informs her actions across large chunks of her life. When the dishonest person lies, then, it is typically neither an incident nor an accident, but expresses a deep part of who she is as a person.

What does it mean to say that vices are *motivated* character defects? In paradigm cases, vicious traits of character dispose someone to act from bad motives because that person has cultivated a wrong orientation towards the good. Thus, Foot (1978) writes that vices involve "false values" (p. 12). Take, for example, the vice of cowardice. If someone can save a person from a burning building at acceptable risk to themselves, then it is *prima facie* cowardly not to do so even if they experience fear. Whether it in fact *is* cowardly, however, will turn in large part on the values that motivate that person. A doctor's decision not to rescue someone from a burning building, for instance, may not be vicious if it is not grounded in a cowardly motivation, but rather in their professional obligations vis-à-vis their patients. This is arguably the case if the doctor is, say, on their way to an urgent life-saving surgery and chooses to prioritize the well-being of patients entrusted to their care. But while the values of a doctor who defers to their professional obligations may be in the right place, the vicious person's values typically fail to capture what is of true moral importance. Rather, the vicious person may be too concerned with, say, wealth, power, and beauty, or they may be overly attached to sensual pleasures. When a cowardly person fails to rescue someone at acceptable

risk to themselves, then, "there is an element of false judgment" (Foot, ibid.) about their failure, for instance because they are too attached to their *own* well-being, or, if they are vain, because they do not want to ruin their good shoes.

Whereas motivation-based conceptions of virtue concern praiseworthy character traits, motivation-based conceptions of vice focus on traits that are blameworthy. An interesting asymmetry between virtue and vice now suggests itself, for, as Crerar (2018) points out, "[b]lame's catchment area" (p. 763) appears to be broader than that of praise. The notion that vices involve false values is a direct application of the inversion thesis, but we routinely blame persons for traits that do not involve false values whatsoever. Consider the range of vices that manifest themselves through such implicit biases as racist or sexist prejudices (Brownstein & Saul, 2016a, 2016b). While there undeniably exist persons who unapologetically embrace racist or sexist values, there is a rich body of evidence that suggests even those who avowedly express egalitarian ideals can exhibit racist or sexist tendencies (see, e.g., Dovidio, Gaertner, & Pearson, 2016 for an overview). Though implicit biases are the subject of much debate, there is an emerging consensus that they can be blameworthy even when they do not align with the consciously adopted values of the biased person (Holroyd, 2012; Sie & van Voorst Vader-Bours, 2016; Washington & Kelly, 2016). In other words, someone's character can be blameworthy if it disposes them to e.g. sexist or racist behavior even if that person's values are properly oriented towards the good.

While it seems plausible that having a praiseworthy character requires a proper orientation towards the good, then, it seems less plausible that a blameworthy character requires an orientation *away* from the good (e.g., towards "false values"). Though adopting values that oppose the good may be a *sufficient* condition for vice on this motivationally demanding conception, it is not a necessary condition. This is shown also by the existence of such vices as *sloth* and *negligence*. These are both deficiencies in motivation: the sloth is *un*motivated, disposed instead towards laziness; and the motivation of a negligent person is neglectful, failing to meet a minimum standard of care. But both the vice of sloth and the vice of negligence are compatible with caring about ends that are themselves morally good. Someone who is negligent, for instance, may value the well-being of others, but through sheer incompetence or thoughtlessness cultivate well-entrenched habits that frequently endanger those around her. When this incompetence or thoughtlessness is culpable – e.g., when she could, and should, have done better – the negligent person is entirely blameworthy for acquiring bad habits and dispositions, though the *values* that motivate her are not blameworthy.

The fact that having a vicious motivation need not involve an orientation towards false values suggests that the inversion thesis is wrong in the case of virtue ethics as well. However, we can construct a motivation-based conception of vice even if we abandon the idea that having a vicious motivation requires adopting morally bad values. While the fact that we blame persons for having a vicious motivation suggests that motivation-based vices are *voluntary* traits, we may exercise a degree of control not just over our values, but also over the other attitudes, habits, and desires that drive us. More specifically, we can often exercise a degree of control over how these attitudes, habits, and desires become entrenched.

In order to see this, it is worth considering how advocates of a motivation-based conception of vice believe these vices are acquired. Insofar as vices are voluntary traits, Foot, Hursthouse, and Zagzebski distinguish vices from natural traits in the same way that they distinguish virtues from natural traits. Foot (1978), for instance, takes great pains to distinguish vices from such innate "pathological fears" (p. 12) as claustrophobia. Although the claustrophobic person resembles the coward – perceiving danger where there is none, unable to control their fears, and so on – we do not blame the claustrophobic person if their fearful disposition is entirely a product of bad constitutive luck, fully the product of influences outside of their control. At least at birth, claustrophobia is not a vice because it is not a *motivated* character defect; that is, it does not reflect the claustrophobic person's voluntary choices. Claustrophobia may turn into a vice, however, if the claustrophobic person has the option of mitigating their fears but fails to do so.

According to advocates of motivation-based conceptions of vice, we acquire a vicious motivation in the same way that we acquire a virtuous one, namely through practice and habituation. Indeed, Zagzebski (1996) characterizes our moral development as providing similar opportunities for developing virtues *and* vices. "At some point in a person's development," she writes, "there must be a realistic possibility that she will develop a certain vice rather than the associated virtue" (p. 105). We develop both vices and virtues over the course of extended interactions with our environment, and it is at least partly up to us what lessons we take from these interactions. An honest person, for instance, may have taken situations calling for honesty as an opportunity to practice being responsible with the truth, while a dishonest person may have taken those same situations as opportunities to test how many lies they can get away with. Either way, both vices and virtues gradually become entrenched as the result of choices we make over the course of our lifetime. Whereas the virtuous person deserves credit for getting things right, however, the vicious person deserves

blame for getting things wrong – not just once or twice, but enough times for her bad habits and attitudes to become entrenched aspects of her identity.

The requirement that we exercise a degree of control over our vices may initially seem to be at odds with counting racist and sexist biases amongst our vices. After all, these biases resemble natural traits in that we have seldom actively cultivated them. Indeed, acquiring implicit biases is often a case of bad moral luck in its own right, as when we passively absorb the prejudices that are prevalent in the communities we grow up in (Begby, 2013; Fricker, 2007). To make matters worse, these prejudices often operate subconsciously, outside of our direct control. While implicit biases will not count as vices on a motivationally demanding conception if they are truly involuntary, however, the developmental picture sketched above helps explain why in other cases they do.

More concretely, implicit biases *do* count as vices if they have become entrenched as the result of choices we have made and if we had "a realistic possibility" (Zagzebski, 1996, p. 105) to counteract them instead. These conditions are plausibly met if we have (or *should* have) evidence that we are biased in some pernicious way and choose not to act on that evidence. Nomy Arpaly (2003) illustrates this by introducing the example of Solomon, "a boy who lives in a small, isolated farming community" and "believes that women are not half as competent as men when it comes to abstract thinking" (p. 103).[21] Solomon's stereotype is not culpable, Arpaly claims, if it is supported by the evidence available to him (e.g., if none of the women in his community engage or display competence in abstract thinking). However, that changes once Solomon has access to evidence that disconfirms his stereotype. In Arpaly's example, Solomon enrolls in a university where he encounters plenty of women who are great abstract thinkers. If Solomon refuses to revise his stereotype in light of this evidence, Arpaly claims, then he *is* blameworthy for his prejudiced attitude. Indeed, to the extent that Solomon continues to act on his bias against women, that bias will become further entrenched in a way that *does* reflect Solomon's agency. After all, at this point he has the option to correct his prejudiced attitude against women.

Natural traits may come to count as vices for the same reason. Though someone with the bad fortune of having an irascible temperament is not blameworthy for being quick-tempered at birth, that changes when they have an opportunity to control their irritability but fail to do so. Thus, if their bad temper becomes entrenched because they allow it to fester, it counts as a motivation-based vice as well.

---

[21] For a detailed analysis of Arpaly's example, see Fricker (2007).

The developmental picture that underlies motivation-based conceptions of virtue and vice also explains why not every trait that falls short of virtue is yet a vice. At certain stages in our life, neither virtuous nor vicious dispositions will be entrenched aspects of our personalities, though we may be well on our way to cultivating either. A child, for instance, is neither virtuous nor vicious because children are both too young to acquire the moral wisdom of a virtuous person and too young to be blameworthy for bad habits and dispositions. Similarly, an adult striving to be virtuous may regularly act in accordance with virtue – i.e., *do the right thing* – without having fully internalized the values and dispositions of a virtuous person. While that person does not count as fully virtuous in that her character traits have not yet settled, she certainly does not qualify as vicious either, and she is due more praise for her character than someone whose values and dispositions prevent her from doing the right thing. By the same token, an adult who reliably acts in accordance with vice – i.e., *does the wrong thing* – may not be vicious, though she will certainly not be virtuous either. This is plausibly the predicament of the biased person who is truly unable to control their prejudice, or of the biased person who has just realized their prejudice and is trying to change but has not yet managed to do so.

All of this shows that both virtuous and vicious motivations come in varying degrees: the virtuous person meets the highest standards for praiseworthiness; the vicious person meets a threshold level of blameworthiness; and the space between is occupied by persons who fit neither category. While the inversion thesis is wrong in suggesting that motivation-based vices involve false values, then, other aspects of motivation-based conceptions of vice mirror motivation-based conceptions of virtue quite nicely. Like motivation-based virtues, motivation-based vices come in degrees; like virtues, these vices are deep and stable qualities of a person that reliably produce a certain type of action; like virtues, these deep and stable qualities are acquired as a result of voluntary choices we make over the course of extended interactions with our environment and thus reflect our agency; and like virtues, the moral status of these vices is intrinsically bound up with the quality of our motivation.

1.4.2 Ends-based Conceptions of Vice

Though many of our vices are characterized by motivational flaws, we sometimes characterize vices in terms of their ends or consequences as well. On ends-based conceptions of vice, vices are any and all character traits that promote morally bad ends or consequences.

Driver (2001) develops an ends-based conception of vice that exactly mirrors her conception of virtue. According to Driver, then, vices are character traits that "normally or systematically [lead] to bad effects" (p. 88). For Driver, the consequences of our character dispositions determine whether they are vicious regardless of what these dispositions reveal about our motivation. On this view, such traits as cowardice, dishonesty, and cruelty are vices because they produce more bad outcomes than good ones, *not* because having a cowardly, dishonest, or cruel personality reveals a blameworthy psychology.

Much like ends-based conceptions of virtue, ends-based conceptions of vice capture the intuition that the moral worth of our character is at least partly a function of the ends and consequences our character traits promote. Indeed, when we evaluate someone's character, we not only look at their motives, but also at the track record produced by their actions. If we find that someone is consistently disposed to bring about bad consequences despite having a virtuous motivation, we may evaluate their character as bad simply in virtue of its pernicious influence on the world. This seems to suggest that factors *external* to our psychology may have normative significance for our character traits as well.

Though the inversion thesis did not cause serious problems for motivation-based conceptions of vice, applying the thesis to ends-based conceptions leads to two counterintuitive results. First, Driver's account elides the space *between* vice and virtue. Only those character traits that systematically produce *as many* good as bad effects are evaluatively neutral. However, those entirely neutral traits are likely few and far between, and this suggests that some traits count as vicious even though they only produce marginally more bad outcomes than good ones (and vice versa). An upshot of this is that someone's character may count as vicious while it is only a few good consequences away from being virtuous. That seems like the wrong result: we typically think of vice and virtue as being on opposite sides of a spectrum that includes a fairly substantial neutral terrain (e.g., in which one is *neither* virtuous *nor* vicious). One possible way out of this conundrum would be to argue that ends-based conceptions of virtue and vice admit of varying degrees too.

A second problem with Driver's account of vice is that it seems to entail that certain children and individuals with (cognitive) disabilities qualify as vicious. Someone's severe cognitive impairments, for instance, may systematically dispose that person towards producing morally bad outcomes through no fault of their own. But to call someone with cognitive disabilities *vicious* on account of their disability seems wrong: after all, we typically reserve that predicate for individuals who have had a hand in their morally deficient character. Driver attempts to take the sting out of this concern by distinguishing between a trait's *being*

bad and that trait's *being blameworthy*. According to Driver, then, imputing a certain vice to someone is simply to point out that they possess a trait that systematically leads to bad effects. This gives that person a strong reason to try and change that character trait if they can, but does not imply that they are blameworthy for having the trait in the first place. However, insofar as Driver is committed to the claim that vices are morally bad, it is not clear that this argumentative move placates the worry. Indeed, the result that the character of disabled persons is morally bad if their disability disposes them towards producing morally bad outcomes seems problematic enough even if we abandon the idea that disabled persons are for that reason blameworthy.

In view of these problems, ends-based conceptions of vice would benefit from further research addressing the exact conditions under which character traits that produce morally bad outcomes qualify as fully-fledged vices. The counterintuitive conclusion that disabilities count as vices can be avoided, for instance, if we define "character" such that disabilities are not part of our character. Indeed, Driver (2001) simply defines character traits as "complex *psychological* disposition[s …] to feel, behave, and/or act" (p. 68) in a characteristic way, which seems to entail that psychological disabilities can be character traits. But if we instead define character traits as dispositions that somehow reflect our *agency* and/or *voluntary choice*s, most psychological disabilities likely do not count. The resulting conception of vice would be a mix between motivation- and ends-based conceptions: a *voluntary* trait that systematically produces bad outcomes.

I do not, however, pursue the project of constructing an ends-based conception of vice further. For present purposes, it suffices to note that just as we can be morally excellent in various ways, we can be morally *deficient* in various ways. This seems to speak in favor of embracing pluralism about *vice* as well. Motivation- and ends-based conceptions of vice each track a different, but equally valid intuition about moral worth. Accordingly, there is space in our moral vocabulary for both conceptions. The former gives name to a deficiency in *motivation*, whereas the latter gives name to a deficiency in our *impact on the world*.

1.5 Conclusion

As we have seen, the concepts of vice, virtue, and character in general play an important role in each of the dominant moral traditions. Across these traditions, the value of these concepts lies in their ability to represent the moral worth of an agent in a particular way. In the virtue

theoretical literature, it is common to distinguish between roughly two conceptions of virtue and two corresponding conceptions of vice. On motivation-based conceptions, virtues are acquired character traits that robustly dispose someone to act from a virtuous motivation, and vices are acquired character traits that robustly dispose someone to act from a deficient motivation. On ends-based conceptions, virtues are character traits that systematically produce more good effects than bad ones, and the other way around for vices. As Battaly (2014) argues, these conceptions are not mutually exclusive insofar as they capture different moral intuitions. Instead, they circumscribe a family of different but related conceptions of human excellence and deficiency. In what follows, then, I will not seek to vindicate either consequence- or motivation-based conceptions. Rather, I will seek to find out whether and how we can apply each of these conceptions to *group* character.

Establishing that we can apply such aretaic concepts as virtue and vice to groups will not be easy. Though there is a burgeoning literature on group virtue and vice (Beggs, 2003; Fricker, 2010, 2013, 2020; Lahroodi, 2007), the majority of virtue theorists – and certainly the ones we have considered thus far – conceive of vices and virtues as features of our individual psychologies. We have seen, for instance, that McDowell (1979) explicitly conceives of virtues as a "perceptual capacity" (p. 332). Hursthouse (2016) maintains that virtue involves human dispositions to "notice, expect, value, feel, desire, choose, act, and react in certain characteristic ways" (para. 7). Though Driver's (2001) account of virtue and vice is one of the least demanding accounts in the literature, she does identify one important constraint: that virtues and vices are "qualit[ies] of the *mind*" (p. 107, emphasis in original). But how can we ascribe complex psychologies involving such deeply subjective mental states as desires, perceptions, and values to groups? Attributing these traits to groups presupposes the provocative idea that groups of individuals can exhibit subjectivity in their own right. If this entails a commitment to the existence of group minds, minds that exist over and above the cognitive faculties of individual agents, we may worry that collectivist accounts of vice and virtue are committed to "mysterious" and "incoherent" (p. 404) ontological claims (Searle, 1990). Arguing for a virtue ethical approach to evaluating group conduct thus requires that we make palatable the idea that we can attribute these traits to group agents. I take up that challenge in the next chapter.

Chapter 2 – The Idea of Group Character

2.1 Introduction

In our everyday linguistic practices, we frequently evaluate the conduct of groups using the language of virtue and vice. We accuse the American aircraft manufacturer Boeing of "arrogance," find the London Metropolitan police service "institutionally racist," recognize fire brigades for their bravery, and we sometimes claim that executive boards are open-minded.[22] There are several things we could be doing when we say things like these. In attributing a virtue to a group, we could be using a shorthand for talk of *individual* character traits. If this is what we are doing, our claim that e.g. a fire brigade is brave amounts to nothing more than the claim that a sufficient number of its individual members are brave. But we could also be making a stronger claim. At least sometimes, it appears that we are using the language of virtue and vice to appraise the conduct of groups *as subjects in their own right*.

Those philosophers who are drawn to the idea that social groups can exhibit virtues or vices as subjects in themselves usually motivate their position by pointing to so-called "divergence arguments."[23] These arguments purport to show that virtues or vices realized at the group level can come apart from the character traits of individual group members in interesting ways. To account for this divergence, one could draw on various theories of group agency. One such theory that is often used to model group virtues and vices is Margaret Gilbert's (1989, 2013) *plural subject theory*. On plural subject accounts of collective virtue and vice, some groups form a Gilbertian plural subject that is virtuous or vicious in its own right (Fricker, 2010, 2013, 2020; Lahroodi, 2007).

While these plural subject accounts offer a promising point of departure for my discussion of group character, the first part of this chapter argues that plural subject approaches suffer from two important limitations. The first limitation restricts the applicability of plural subject accounts to motivation-based conceptions of virtue and vice.

---

[22] See, e.g., Shepardson (2020) and Laker, Cobb, and Trehan (2021) on Boeing, and the Macpherson Report (1999) on institutional racism within the London Metropolitan Police.
[23] I borrow the term "divergence argument" (p. 343) from Jennifer Lackey's (2016) discussion of *group justified belief*.

Since plural subjects are instantiated when two or more individuals jointly commit to upholding a certain feature as a body, plural virtues of a motivation-based kind can be willed into existence simply by committing to virtuous motives. The consensus in much of contemporary virtue theory, however, is that motivation-based virtues are genuine achievements, acquired excellences that are deeply praiseworthy at least partly because they "develop over a long period of time" (Zagzebski, 1996, p. 128). Accordingly, plural subject accounts run the risk of making what should be a significant achievement – the acquisition of a virtuous motivation (or indeed a vicious one) – too easy to realize.

The second limitation concerns the *stability* of plural virtues and vices. Since their constitutive joint commitments can be rescinded at will (subject, as we will see, to the approval of all parties involved), plural virtues and vices often fail to achieve the level of stability characteristic of robust character traits. Just as plural subject accounts risk making collective virtues and vices too easy to realize, that is, plural subject accounts risk making these traits too easy to lose. Plural subject accounts therefore have difficulties accommodating a central feature of most contemporary accounts of virtue and vice.

In view of these limitations, there are strong reasons for considering alternative accounts of collective virtue and vice. In the second part of this chapter, I argue that we should embrace a *functionalist account* of group agency instead (see, e.g., List & Pettit, 2011). On the novel, functionalist account of collective virtue and vice I develop, collective virtues and vices arise when organizations or other collective actors are organized so to as to function *as* virtuous or vicious agents. The details of this account will depend on one's preferred conception of virtue and vice. As we have seen in chapter one, motivation-based conceptions of virtue and vice impose more demanding requirements on the psychology of moral agents than ends-based ones, including requirements on the values and motives that drive the agent. Accordingly, the functionalist account developed here imposes more stringent conditions on virtuous or vicious functioning in a motivationally demanding way than it does on virtuous or vicious functioning in ends-based respects. We will see, however, that all of the features standardly associated with virtues and vices – including motives and values – can be modelled in functionalist terms. The resulting view of collective virtue and vice can thus model a broad range of collective character traits and is preferable to existing alternatives.

This chapter is structured as follows. Section two briefly summarizes the salient details from chapter one, reminding us of the challenges that will face us when we apply the concepts of virtue and vice to collectives. Once these concepts are in focus, I discuss how they may apply to groups in section three. I present my critique of plural subject accounts of

collective virtue and vice in section four. This critique sets the stage for my central argument – that we should embrace a *functionalist* account of collective virtue and vice – in section five. While the preceding sections show that a functionalist account is preferable to existing accounts of collective virtue and vice, I discuss the advantages of a functionalist account for future research in section six. We will see that a functionalist account of collective virtue and vice not only has considerable empirical purchase, but also meshes well with literature from the organizational sciences. Section seven concludes.

2.2 The Nature of Virtue and Vice

Virtues are character traits that make us excellent. Vices are character traits that render us deficient. As we have seen in the previous chapter, contemporary virtue theorists commonly distinguish between roughly two conceptions of virtue and two corresponding conceptions of vice: *ends-based* conceptions and *motivation-based* conceptions (Battaly, 2010, 2014).[24] I briefly discuss each in turn.

On an ends-based conception, virtues are traits that reliably help us attain good ends or consequences. Accordingly, one must show that groups are capable of manifesting character dispositions that reliably advance good ends or produce good consequences in order to apply an ends-based conception of virtue to collectives. The main contemporary proponent of an ends-based conception of virtue is Julia Driver (2001), but we also find this conception at work in the reliabilist tradition within virtue epistemology. For reliabilists like Ernest Sosa (2007) and John Greco (2010), epistemic virtues are valuable traits because they reliably help us attain such epistemic goods as knowledge, wisdom, and understanding. Reliabilist vices stand in the way of attaining these goods. The most influential account of such vices is offered by Quassim Cassam (2019). According to Cassam, epistemic vices are those traits that reliably "obstruct the acquisition, retention, and transmission of knowledge" (p. 7).

On a motivation-based conception, the excellent nature of virtue does not just pertain to the ends of virtue, but is intrinsically bound up with the quality of our motives. Showing that groups can manifest virtues of this kind thus entails showing that they can act from a virtuous motivation. Contemporary champions of a motivation-based conception of virtue and vice include Philippa Foot (1978, 2001), Rosalind Hursthouse (1999), and Michael Slote

---

[24] Miranda Fricker (2010) follows Julia Annas (2010) in distinguishing between a Stoic or "skill-based" conception of virtue and an Aristotelian or "motive-based" conception of virtue. Roughly, the Stoic conception is an instance of what Battaly calls an ends-based conception.

(2001). Motivation-based conceptions of virtue and vice are also prominent within the responsibilist tradition in virtue epistemology. Linda Zagzebski (1996), for instance, argues that epistemic virtues not only help us achieve epistemic goods, but further involve a proper *orientation* towards these goods. In the domain of epistemic vice, Jason Baehr (2010) argues that at least some of our intellectual vices involve a wilful "opposition to knowledge as such" (p. 204), illustrating that even our epistemic motives can be deficient (see also Tanesini, 2018).

Although the proponents of motivation- and ends-based conceptions of virtue and vice usually take themselves to be advancing mutually incompatible views, I argued in the previous chapter that we have reason to be optimistic about the prospects of a pluralist account of virtue and vice that encompasses both conceptions (see, e.g., Battaly, 2015; Swanton, 2003; Wood & Roberts, 2007). Accordingly, I will aim to develop an account of collective virtue and vice that can accommodate *both* types of excellence and deficiency. While I take ends-based conceptions of virtue and vice to be important additions to our virtue theoretical vocabulary, the discussion below nevertheless concentrates on virtues and vices of a motivation-based kind. This is for the purely methodological reason that a successful ends-based account of group virtue and vice *falls out of* a successful motivation-based account. Both conceptions of virtue involve reliably performing well with respect to morally or epistemically valuable ends (or consequences), but motivation-based conceptions are more demanding in further requiring good motives. The inverse holds for vices. Hence, if we can show that groups can instantiate motivation-based virtues or vices, we have also shown that they can exhibit ends-based virtues or vices. As we will see below, this is one of the reasons I will argue that a functionalist account of collective virtue and vice is preferable to plural subject accounts. While plural subject accounts have difficulties modeling important features of motivation-based virtues and vices, a functionalist account can model excellences and deficiencies of both kinds.

2.3 Collective Virtues, Collective Vices

Historically, virtue theorists have mostly restricted the concepts of virtue and vice to the character of individual human agents. Thus, Philippa Foot (1978, p. 6) and Alasdair MacIntyre (1981, p. 189) argue that virtues are indispensable for a complete "human life," while Martha Nussbaum (1990) writes that an investigation into the nature of virtue is

typically an investigation into the "human good," restricted to human qualities, and not into "the good of some other being" (p. 66). In recent years, however, the concepts of virtue and vice have found increasingly broader applications. Nussbaum (2006) herself has started using virtue theory to study the flourishing of other species, and the advent of social ontology has seen a number of philosophers applying the concepts of virtue and vice to *group agents* (Beggs, 2003; Byerly & Byerly, 2016; de Bruin, 2015; Fricker, 2010, 2013, 2020; Jones, 2007; Konzelmann Ziv, 2012; Lahroodi, 2007; Sandin, 2007).

The search for an account of collective virtue and vice derives much of its appeal from so-called *divergence arguments*. These arguments usually involve cases that purportedly show how group-level traits can come apart from the traits displayed by individual group members. One of the standard divergence arguments in the literature on group virtue and vice is due to Reza Lahroodi (2007). Lahroodi asks us to imagine a church committee that is collectively narrow-minded about gay rights, though it is entirely composed of individually open-minded members. As private persons, these members routinely take ideas favoring gay rights seriously. When they come together as a committee, however, the group "summarily dismisses" progressive ideas involving gay rights, not deeming them "worthy of discussion, let alone adoption" (p. 287). It is easy to imagine, for instance, that though they are each open-minded individuals, the committee members feel pressured to uphold the church's stance on gay rights when they act in their joint capacity as decision-makers within the church.

In Lahroodi's example, the traits displayed by individual group members diverge from traits instantiated at the group level. While it seems apt to describe the church committee as narrow-minded, it does not seem apt to describe any of its members in that way. This divergence straightforwardly disqualifies a simple *summativist* analysis of group traits as a general account of group features. On a summative analysis, a group G possesses virtue (or vice) V if and only if a sufficient number of its members display V.[25] But here we have a case where we cannot explain the group-level trait of narrow-mindedness by tallying that trait at

---

[25] Anthony Quinton introduces the term "summative" (p. 9) in his (1976). See Gilbert (1989) for a more thorough account of the summativist position. The summativist position roughly corresponds to what Lahroodi calls the "individualist" (p. 281) position and Anita Konzelmann Ziv (2012) calls the "distributive or aggregate" (p. 88) position.

the level of individual group members. Lahroodi concludes that we apparently require a more sophisticated account of how individuals combine to realize group-level virtues.[26]

To account for this divergence, both Lahroodi and Miranda Fricker (2010, 2013, 2020) initially turn to Gilbert's (1989, 2013) *plural subject theory* of group agency. According to Gilbert, two or more individuals can instantiate the plural subject of a certain collective feature by jointly committing to upholding that feature as a body. The members of a hiring committee, for example, can jointly commit to upholding the belief that a woman should fill the position of CEO, thereby making it true that the hiring committee (*qua* plural subject) has that belief. As Fricker emphasizes, this does not entail that the individual committee members share that belief as private individuals. Invoking Christine Korsgaard's (1996) notion of "practical identities" (p. 101), she points out that our commitments often vary according to our membership in different social groups (that is, according to our different practical identities).[27] In their role as private individuals, for example, the committee members may believe that gender should not factor into hiring decisions (or perhaps a bigoted few believe that only men should be considered). But so long as they are guided by the belief that a woman should fill the vacancy in their practical identity *as* members of the hiring committee, we can still ascribe that belief to the group as a plural subject.

Although I will discuss Gilbert's notion of a plural subject in further detail below, it should be relatively easy to see how we can use the concept of a plural subject to construct an account of collective virtue and vice. All that is needed to instantiate a collective virtue or vice is for a group of individuals to jointly commit to a virtuous or vicious trait and for the group to be reliable in upholding their commitment. Applying this analysis to the motivation-based picture of virtue introduced in the previous section, Fricker (2010) offers the following account of collective virtue:

> If … a number of individuals [jointly] commit to a virtuous moral or epistemic
> motive, they thereby constitute themselves as the plural subject of that *collective*
> *motive*. Joint commitment to a motive should be understood here as shorthand for a
> joint commitment to achieving the good end of the motive *because* it is good […] Let

---

[26] This is not to say that a summativist analysis is *always* inappropriate. When we commend a fire brigade for their bravery, we may well intend to praise the characters of its individual firefighters. See also Todd T. Jones (2007) on "correlativism" (pp. 442-445).

[27] The notion that our (ethical) commitments may vary according to our practical identities is somewhat controversial. In the literature on integrity, for instance, it is often assumed that we are subject to the same moral codes regardless of what practical roles we occupy (see, e.g., Bigelow & Pargetter, 2007; McFall, 1987). Still, we do often appear to act *as if* our commitments vary according to social roles.

us now add to this group motive the requisite reliability condition; and *voilà*, we have

a collective virtue.[28] (pp. 241-242, emphasis in original)

To meet this reliability condition, Fricker adds, the group's motives must be *counterfactually stable*. We do not ascribe virtues on the basis of one-off good performances. Rather, virtues *consistently* involve good performances in relevant situations. She therefore clarifies in her more recent work that collective virtues usually involve the joint commitment to an "ethos" of robust value commitments that dispose the group to act from virtuous motives across a wide range of appropriate circumstances (Fricker, 2013, p. 1329; 2021, p. 90).[29] Vices, by contrast, involve either the joint commitment to an ethos that disposes the group to act from deficient motives or the group's culpable failure to implement its virtuous commitments.

Returning to Lahroodi's example, the church committee counts as collectively narrow-minded on this view if its members have jointly committed to an ethos that disposes the committee to dismiss out of hand ideas favoring gay rights. As we have seen, this does not require that the committee members are narrow-minded as private individuals. It is perfectly consistent with this view that the committee members are disposed to reject gay rights *only* in their role as members on the committee, and not in the context of the practical identities that govern the rest of their lives. The same holds true for collective virtues. For instance, a team of journalists can instantiate the virtue of diligence by jointly committing to an ethos of diligence and reliably living up to that commitment even if it is entirely composed of individuals who are nothing but careless in their private lives. Since in cases like these it is only in their capacity as group members that the persons involved display virtuous or vicious behavior, Fricker (2010) concludes that "there is no level of group-*independent* features to which the higher-level [virtues or vices] can be reduced" (pp. 238-9). We have thus apparently vindicated a *non-summativist* account of collective virtue and vice according to which groups can be the locus of virtues and vices in their own right – that is, *qua* group.

In the section that follows, I will argue that joint commitments are often an inadequate basis for the robust value commitments that Fricker claims are at the heart of collective virtues and vices. Before we move on, however, it is worth pausing on another potential

---

[28] Fricker (2010) also offers a plural subject analysis for what she calls a "skill-based" conception of virtue (a species of *ends-based* conceptions), but later argues that this analysis is "otiose" because we already have a rich vocabulary of "performance-oriented terms of assessment" for collectives (Fricker, 2020). Although I limit my discussion to motive-based conceptions of virtue for methodological reasons, there is nevertheless an important point in maintaining our ends-based vocabulary. This is because virtue (and vice) terms such as *greedy* or *generous* are typically more *descriptive* than such performance-oriented terms as e.g. *accurate* and *efficient* and thus tell us more about *how* an institution performs.
[29] See Byerly and Byerly (2016) for another dispositional account of collective virtue.

problem with a plural subject account. Lahroodi (2007) ultimately rejects a plural subject analysis of collective virtue because he worries that the requirement that group members (jointly) *commit* to a virtue is too strong. For Gilbert, joint commitments involve an *awareness* of what it is that is committed to, but many virtues (and indeed vices) are such that we need not be aware of having them.[30] I agree with Fricker (2010) that Lahroodi's concerns are misplaced: the parties to a virtuous or vicious plural subject may simply not know that the motive they have jointly committed to *counts* as virtuous or vicious. It is unlikely, for example, that Lahroodi's church committee has explicitly embraced narrow-minded motives under the description of being *narrow-minded*. Indeed, if the committee routinely rejects the importance of gay rights out of hand, it may qualify as narrow-minded even if the committee itself conceives of this tendency as resulting from a noble commitment to church dogma. The claim that plural virtue and vice requires a commitment to virtuous or vicious traits, then, is not tantamount to the claim that the parties involved must conceive of their commitments *as* virtuous or vicious.

2.4 The Limitations of a Plural Subject Account

Fricker's account offers a compelling analysis of collective virtue and vice. Among its significant advantages is that it draws on decades worth of literature on Gilbert's plural subject theory, which has emerged as one of the dominant contemporary theories of group agency. In drawing on plural subject theory, however, Fricker also incurs the theoretical limitations of a Gilbertian framework. As I will argue below, this restricts the applicability of a plural subject account of collective virtue and vice in two ways.[31] The first limitation restricts the applicability of a plural subject approach to motivation-based conceptions of virtue and vice. Indeed, we will see that motivation-based conceptions typically view the acquisition of a virtuous or vicious motivation as a significant achievement that is difficult to accomplish. Since plural subject accounts hold that plural virtues and vices are simply instantiated by jointly *committing* to virtuous or vicious motives (provided that these are appropriately stable), I will argue that these accounts risk making collective virtues and vices too easy to realize. This severely diminishes the evaluative connotations usually associated

---

[30] See, e.g., Driver's (2001) "virtues of ignorance" (pp. 16-41) and Cassam's (2019) "stealthy vices" (pp. 144-66).
[31] For other critical discussions of Fricker's account, see Byerly and Byerly (2016), Cordell (2017), and Konzelmann Ziv (2012).

with virtues and vices. The second limitation of a plural subject approach concerns the stability characteristic of robust character traits. Because joint commitments can be rescinded at will – provided that they are *jointly* rescinded by all parties involved – I will argue that plural virtues and vices often lack the stability required of genuine virtues and vices.

2.4.1 Acquiring Motivation-based Virtues and Vices is Hard

The first problem with a plural subject analysis of motivation-based virtues and vices is that it threatens to reduce what should be a significant moral or epistemic achievement – the acquisition of an entrenched disposition to act or cognize in a particular way – to the mere act of (jointly) committing to a particular motive or ethos. To illustrate, consider the following example:

> **Honest Tobacco** – Healthy Tobacco is a tobacco company run by a board of individually greedy and dishonest executives. For decades, these executives have been "merchants of doubt" (Oreskes & Conway, 2010), peddling misleading evidence and half-truths in order to discredit the significant body of research showing the health risks of smoking – all of this to shore up the company's profits. Slowly, however, the executives realize that the company's reputation as a morally bankrupt organization is catching up to them, hurting the profitability of the company. After consulting a PR firm, the *Healthy Tobacco* executives decide to rebrand. Their company will henceforth be named *Honest Tobacco*, and from now on the company will be completely open and transparent about the health risks of smoking. However, the PR firm has warned them that virtue signaling for purely financial motives is likely to backfire. The executives therefore jointly commit the company to a policy of honesty *for its own sake* – that is, for the reasons that make honesty a valuable trait. As it turns out, the *Honest Tobacco* executives still lack this commitment to honesty as private individuals. Every executive gives her personal assent to the joint commitment out of purely financial motives, knowing that honesty is good for business.[32]

Let us suppose that *Honest Tobacco* is a highly efficient company, such that it is reliable in achieving whatever it commits itself to. Let us further suppose that its value commitments immediately dispose the company to act from honest motives and are immediately reflected in business practice. On the analysis offered by Fricker, *Honest Tobacco* now counts as a plural

---

[32] This case is inspired by Lackey's (2016) work on group (epistemic) justification.

subject of the motivation-based virtue of *honesty*: its executives have jointly committed to upholding (a) honest motives, for (b) honest reasons, and are (c) reliable in acting from honest motives for honest reasons, at least in their joint capacity as board members. That seems too easy.

One reason to be immediately suspicious of a plural subject analysis is that it elides the hard moral or epistemic work (e.g., time and practice) that typically goes into the acquisition of an entrenched character disposition. The consensus view in the literature on motivation-based virtues and vices is that cultivating virtuous or vicious traits is difficult (Hursthouse, 1999) and cannot be accomplished "with the flip of a switch" (Zagzebski, 1996, p. 196). One cannot simply *decide* to become virtuous (or vicious) because a single act of will is "causally insufficient" (ibid., p. 120) to produce the deep states of character that virtues and vices are. As Hursthouse (1999) puts it:

> A change in such character traits is a profound change, one that goes, as we say, 'all the way down.' […] It is certainly not a change that one can just decide to bring about oneself overnight, as one might decide to break the habit of a lifetime and cease to have coffee for breakfast. (p. 12)

We tend to think of virtues or vices as settled states of character that are virtually immune from sudden changes because they are deeply entrenched aspects of our personality; under ordinary circumstances, developing these dispositions requires significant amounts of time, effort, and practice.[33] In the rare cases that we *do* observe a sudden change in character, moreover, this is not the result of deliberate choice but requires "special explanations" such as "religious conversion," "brain damage," or "an experience that changes the person's whole outlook on life" (ibid.).

As we have seen in the previous chapter, the standard Aristotelian story, which Hursthouse and Zagzebski endorse, is that virtues are acquired through *habituation*. One becomes virtuous by developing a habit of acting virtuously (e.g., for virtuous reasons) until one has internalized the moral or epistemic outlook encompassed by these reasons and achieved a deep understanding of what acting virtuously entails. Cultivating the virtue of justice, on this view, requires that one practices the qualities of being fair and impartial until, with time and practice, one acquires a sensitivity to the demands of justice across different circumstances and the habit of acting justly becomes second nature. This is not just a point

---

[33] Aristotle (ca. 350 B.C.E./2020) describes the virtuous person's character as "firm and unchangeable" (1105a-b).

about the *etiology* of virtues, but also about the depth of our praise for virtues. Since virtues are produced by habituation, virtuous states of character are indicative of a sustained *history* of acting virtuously; hence our praise for the virtuous person typically extends well into the past and goes "far beyond the moment of action" (Hursthouse, 1999, p. 123). Roughly the inverse holds for vices. Unlike the virtuous person, the vicious person has cultivated poor habits, and is deeply blameworthy for the history of poor judgments indicated by their vicious state of character.

On a motivation-based conception of virtue and vice, then, these traits cannot simply be willed into existence. As the case of *Honest Tobacco* illustrates, this is precisely where plural subject accounts of collective virtue and vice fall short. Making virtues and vices a matter of joint commitment, plural subject accounts allow that these traits can be produced by a single – albeit joint – act of will; namely, the joint act of *committing* to virtuous or vicious motives. But the dishonest executives who jointly commit to an ethos of honesty are like the thief who decides that "from *now* on, I'll be an honest person." That is to say, barring special explanations, they instantiate no virtuous character trait whatsoever – at least not *yet*. Though the decision to commit to virtuous motives may give one a reason to act virtuously, having a virtuous *character trait* requires a much deeper commitment to virtuous motives, namely one that is anchored in one's entrenched moral or epistemic outlook and is typically evidence of one's virtuous track record. Other things being equal, a single act of will is an insufficient basis for bringing about this level of entrenchment.

Note, by the way, that Fricker's reliability condition does not alleviate these concerns. Even if *Honest Tobacco* is reliable in implementing its commitment to honest motives, such that its commitment to honesty is temporally and counterfactually stable, it will not count as virtuous in the sense intended by e.g. Hursthouse and Zagzebski before these motives have become robustly entrenched over time. In the (neo)-Aristotelian tradition, it simply *is* part of being virtuous that one's virtuous character traits reflect more than a single act of will (and indeed, typically reflect an extended trajectory of self-cultivation).[34]

At this point, critics may object that I have taken the analogy between plural subjects and individual human agents too far. Hursthouse and Zagzebski appeal to Aristotle's account of virtue acquisition at least partly because it is mainly through habituation that they believe human agents can come to possess voluntary traits robust enough to count as virtues or vices,

---

[34] Nor is this purely a feature of the (neo-)Aristotelian tradition. Non-Western virtue traditions including Confucianism and Buddhism also emphasize the importance of self-cultivation, a process that requires significant time and effort (see, e.g., Vallor, 2016, pp. 61-75 for an overview).

but it seems possible that plural subjects can instantiate features akin to entrenched states of character without going through lengthy processes of habituation. Peter French (1995), for instance, contends that group agents such as "[corporations] will probably prove more capable of consistent and dependable ethical behavior than humans" (p. 80) in virtue of their constitutive structures.[35] If *Honest Tobacco* is reliable in upholding its honest commitments because of its strong institutional structure, perhaps the company really *does* have a robustly honest character from the moment its executives jointly decide on a policy of honesty. In that case, plural subject accounts would succeed in modeling collective virtues and vices after all.

Although I will argue that joint commitments are an inadequate basis for a robust institutional character below, there is some merit to this objection. To be sure, we should be wary of making category mistakes. Throughout her work, Gilbert has insisted that plural subjects and human agents exhibit a different *kind* of agency. Joint commitments, Gilbert maintains, bind the parties to a plural subject to act *as if* they collectively formed a single agent. In light of this, it would be a mistake to hold that e.g. plural beliefs are *literal* beliefs (i.e., mental states had by a "group mind"); rather, they are a kind of *as-if* beliefs that arise in virtue of member commitments to "emulate" (Gilbert, 2013, p. 8) an agent that holds these beliefs. It would likewise be mistaken to attribute *literal* virtues and vices to plural subjects; rather, plural virtues and vices are *simulated* qualities that guide the behavior of group members in their capacity *as* members. While we will see in section five that collective virtues and vices are often produced in the same way as their individual counterparts, it at least seems like a *conceptual* possibility that these simulated qualities can be instantiated without a causal history of good or bad judgments.

Even if we grant this, however, it does not follow that a joint commitment to virtuous or vicious motives can instantiate the collective equivalent of individual virtues or vices. To see this, recall that the Aristotelian story of virtue acquisition serves not only an etiological point, but also tracks the quality of our praise or blame. Zagzebski (1996) even goes so far as to suggest that the *primary* reason that virtues cannot be produced by a single act of will is that "a single act of will is logically insufficient to transform oneself into a person whose resulting quality is something we would praise in the sense we praise a virtue" (p. 121). This, she argues, is why e.g. a Nozickian "transformation machine" (Nozick, 1974, p. 44) could not produce a virtuous person. Though it is not incoherent to imagine a machine that can

---

[35] Per Sandin (2007) argues that this is why collective virtues and vices are likely immune from the *situationist challenge* to virtue theory. I return to this claim when I discuss the situationist challenge in the next chapter.

transform our character and thus produce superficially virtuous or vicious dispositions, such a machine could never replicate the history of good or bad judgments leading up to the formation of fully-fledged virtues or vices. Accordingly, a Nozickian transformation machine cannot produce traits for which we are responsible in the deep sense that we are responsible for acquiring a virtuous or vicious motivation. Any praise or blame attached to the resulting qualities would therefore be severely diminished.

The same is true of the sudden changes in character imagined by Hursthouse. Even in the rare occurrence that someone's entrenched dispositions suddenly change after a transformative experience, this change does not reflect the agency of that person to the extent of someone who has gradually cultivated a virtuous or vicious motivation over time. Consider someone whose dispositions change after a spontaneous religious conversion. Suppose a vicious criminal is disposed towards immoral behavior at time t_0, has a flash of inspiration and converts to Christianity at time t_1, and is immediately disposed towards moral behavior at times $t_1 - t_n$. While that person is certainly praiseworthy for acting in accordance with virtue from t_1 onwards, their character is not praiseworthy to the extent of a virtuous person (at least, not at the moment of conversion). The virtuous person is deeply praiseworthy precisely because their character reflects the choices they have made over a lifetime of learning. Of course, this is not to say that the converted criminal could not achieve a character that is praiseworthy in this way. With every moral act they perform, the converted criminal will solidify their habit of acting virtuously and deepen their understanding of what acting virtuously requires, until at some point their virtuous disposition deeply reflects their agency as well.

Divorcing the concepts of collective virtue and vice from the time, effort, and practice that goes into the acquisition of a robustly virtuous or vicious motivation, then, comes at the cost of an impoverished sense of praise or blame attached to collective virtues and vices. This is a price one may be willing to pay, but it is not clear that we should. Fricker (2010), for instance, wants to maintain "the normal implications of praise or blame" (p. 240). Indeed, Fricker (2020) claims that talk of collective character is valuable precisely because it tracks how collectives such as corporations and institutions may "genuinely *stand for something*" (p. 91, emphasis in original). But we rarely praise or blame a collective for the values it stands for prior to an established track record. Surely we would not praise *Honest Tobacco* for standing for the virtue of honesty from the outset, even if it counts as an honest organization on the plural subject account under consideration.

Moreover, we do not just withhold praise from organizations like *Honest Tobacco* for the epistemic reason that we lack evidence about their values. Rather, we withhold praise because standing for something *in the way that a virtuous person stands for something* requires that one's value commitments robustly reflect one's agency over time. If having a virtuous motivation is a deep "credit" (Battaly, 2014, p. 54) to someone, this is because they have contributed more than a single act of will to acquiring the values, wisdom, and attitudes characteristic of a virtuous person. If we want to be able to praise or blame collective actors in the same way that we praise or blame a virtuous or vicious individual, then, we should avoid making collective virtues or vices too easy to realize. As I have argued here, making collective virtues or vices a matter of joint commitment will not do, even if we include a reliability condition.

2.4.2 Plural Virtues and Vices are Unstable

The second problem with a plural subject account of collective virtue and vice is that plural virtues and vices often lack the stability required of genuine virtues and vices. Virtue theorists tend to think of virtues and vices not only as traits that are difficult to realize, but also as traits that are relatively immune from sudden changes. Once virtues and vices are entrenched, they dispose their possessor to virtuous or vicious behavior across a wide range of circumstances. Having cultivated the virtue of honesty, for instance, the honest person is not easily tempted into *dis*honesty and can be relied upon even if lying or cheating would be politically or economically expedient.

Joint commitments are an inadequate basis for this kind of stability because they can be rescinded at will (if, that is, the parties involved agree to rescind it). This claim may sound surprising at first. After all, Gilbert insists that joint commitments create strong reciprocal normative ties from which participants cannot unilaterally withdraw without violating social norms. Indeed, these ties are such that the parties involved have "the standing to […] rebuke for non-conformity" (Gilbert, 2013, p. 9).[36] As Fricker (2020) observes, this aspect of joint

---

[36] Konzelmann Ziv (2012) uses this feature of joint commitments to object to Fricker's plural subject approach, arguing that (a) Fricker is committed to understanding virtues as *supererogatory* qualities, but that (b) we cannot rebuke someone for failing to do more than is required of them. Even if we grant (a), (b) rests on a misunderstanding of Gilbertian joint commitments. Gilbert emphasizes (2008, 2013) that the standing to rebuke someone for reneging on a joint commitment is not a *moral* standing. It is the same standing we have to reproach ourselves for failing to uphold a *private* commitment (e.g., the standing I have to reproach myself for failing to work out more despite committing to a healthier lifestyle). When I rebuke my partner for reneging on our joint commitment to a virtuous trait, then, I do

commitments actually *enhances* the stability of joint commitments in some respects, since members that "[prove] their commitment less than appropriately stable by reneging on it […] are properly subject to rebuke" (p. 95). When two individuals have jointly committed to the virtue of diligence and one of them violates their commitment, for example, the other can hold that person to account – "Why are you so careless? That's not what we agreed!" – in order to bring them back in line.

 While the parties to a joint commitment lack the standing to rescind their commitment *unilaterally*, however, there is nothing that stops them from doing so together. Upon finding that the joint commitment to a certain feature no longer serves their interests, the parties to that commitment may *jointly* decide to withdraw from their commitment after which it is no longer binding (Gilbert, 2013, p. 40). It is in this regard that *Honest Tobacco's* commitment to honesty is vulnerable to sudden change. At any point in time, *Honest Tobacco*'s executives may choose to repudiate their joint commitment to honest motives and so dismantle the putatively honest plural subject. Given their greedy and dishonest temperaments (*qua* private persons), moreover, it is likely that they will do exactly this when the financial motive is there. But an agent that simply repudiates the trait of honesty when doing so is financially expedient does not count as having the virtue of honesty at all.[37] It thus seems that joint commitments alone cannot carry the moral (or epistemic) weight we usually associate with attributions of motive-based virtues and vices.

 At this point, I should acknowledge that Fricker agrees that collective virtuous or vicious motives must be appropriately stable. She specifically means for this stability requirement to rule out cases like *Honest Tobacco* as instances of collective virtue: "a purportedly 'green business' will not count as possessing the virtue of being green if the moment it were to meet a conflict between ecological values and profit motive it would ditch the values to pursue the greater profit" (Fricker, 2013, p. 1327). Just like individual virtues and vices, Fricker writes, their collective counterparts must involve "fairly resilient collective value commitment[s]" (ibid.) that persist over time and can withstand countervailing motives. She proposes that we locate these value commitments in an organization's "ethos" of counterfactually stable "collective motivational dispositions and evaluative attitudes"

---

not rebuke them for not doing more than is ethically required of them, I rebuke them for not doing what *is* required of them as a matter of practical rationality, namely, to follow through on their commitments.
[37] One aspect that is often thought to set virtues apart from other valuable traits such as skills is that repudiating the former, but not the latter, counts against possession of the trait in question (see, e.g., Broadie, 1991, p. 89). A skilled piano player who gives up on playing the piano is no less skilled, but a just person who gives up on acting justly is no longer just.

(Fricker, 2020, p. 91). In light of this, it may seem uncharitable to say that Fricker's account does not guarantee the standard of stability required of virtues and vices.

Yet introducing the notion of an organizational ethos only exacerbates the problem. This is because Fricker operationalizes that notion entirely in terms of joint commitments itself:

> It is precisely the bindingness of Gilbert's model that makes it the right one for present purposes, for in order to make sense of the bindingness of ethos we *need* the commitment involved in the joint commitment that creates the plural subject of the ethos. An ethos is precisely not something *pro tem*, but something committed and intersubjectively binding by way of potential apt rebuke. A set of values that one can ditch when it no longer suits is no ethos at all, but mere lip service. (ibid., p. 95)

According to Fricker, then, an ethos brings stability to an organization because the joint commitment *to* that ethos generates an internal pressure to conform. As we have just seen, however, that internal pressure is powerless if the parties to a joint commitment choose to withdraw from their commitment *together,* at which point they lose the standing to rebuke for non-conformity. Just like the joint commitment to a virtuous or vicious motive, the joint commitment to an ethos can be rescinded at will.

Of course, one may object that though it is logically possible to rescind joint commitments at any time, some such commitments seem stronger than others as a matter of contingent fact. Since joint commitments can only be rescinded if all parties agree, they are unlikely to be rescinded, for instance, if the majority of participants care strongly about the commitment in question. Moreover, the commitment to an ethos consisting of virtuous values and dispositions may be exactly the kind of commitment that inspires such loyalty. Other commitments that may not be rescinded so easily are those subject to legally binding documents or contracts such as corporate charters and codes of conduct. In these cases, however, it is important to observe that factors *external* to the joint commitment are what make the organization reliable in upholding its values.

The upshot of this discussion is that the standard of motivational robustness required for virtues and vices simply is not achieved by joint commitments alone. Put differently, there is no *conceptual reason* that prevents joint commitments from being rescinded at any time. Nothing in the logic of Gilbert's model prevents a set of values that is jointly committed to from being "mere lip service." But just as one cannot instantiate a virtue or vice through a single act of will, one should not be able to *repudiate* a virtue or vice through a single act of will. To give an account of collective virtues and vices, then, we need to explain how a high

56

degree of motivational robustness can be realized *independently* of joint commitments. And while we will see that the functionalist account of collective virtue and vice I go on to endorse in the next section also includes a stability requirement that is extraneous to the account of group agency on which it is based, it is not committed to saying that collective virtues or vices can be instantiated or repudiated by a joint act of commitment (i.e., will).

Before we continue, I should emphasize that the limitations of a plural subject account do not restrict the applicability of these accounts to *all* conceptions of virtue and vice. On the ends-based conception of virtue defended by Driver (2001, 2016), for instance, it is neither required that one has virtuous motives nor that one's character traits are inherently stable, so long as these traits systematically produce good outcomes. Accordingly, a plural subject approach seems more than capable of modeling *these* types of excellences. My aim, however, is to develop an account of collective virtue and vice that can model motivation-based virtues and vices as well. In the next section, I argue that a functionalist account of collective virtue and vice provides a promising blueprint.

2.5 A Functionalist Account of Group Character

In the previous section, I argued that Fricker's use of Gilbert's plural subject theory limits her account of collective virtue and vice in two ways. First, plural virtues and vices of a motivation-based kind are too easy to realize. As a consequence, their praise- or blameworthiness is severely diminished. Second, plural virtues and vices often lack the stability required of fully-fledged virtues and vices because they are subject to rescission. Since both these limitations are due to the constitutive role of joint commitments, we should look for an account of collective virtue and vice that does not turn on the presence of such commitments. In the remainder of this chapter, I argue that we should adopt a *functionalist account* instead.

2.5.1 Setting up the Philosophical Challenge

Before we discuss a functionalist account in detail, let us remind ourselves of the relevant philosophical challenges. To do so, it is worth examining a scenario Fricker (2010) does *not* believe involves collective virtue:

> Imagine a debating society, the members of which are thoroughly prejudiced individuals. But their prejudices are all opposed and equally balanced, so that they cancel each other out in debate and the debate overall displays not prejudice but rather neutrality […] Such a debating society, then, is itself non-prejudiced, even while all of its members, and their contributions to the group's activities, are prejudiced. (p. 239)

In this example, a kind of "invisible hand" (ibid.) generates the group-level trait of being non-prejudiced. But this trait does not count as a *virtue*, Fricker argues, because it is a "mere accident of the way the individual's prejudices happen to cancel one another out" (ibid.). Hence, there is no guarantee that the group will continue to be non-prejudiced if its membership changes or if existing members change their views. Virtues are traits that are more counterfactually stable than that. This poses a first constraint on our account of collective virtue, which I will call the *stability requirement*.

While the stability requirement is important, it is not the only constraint on an account of collective virtue. We could easily re-imagine Fricker's example such that it meets the stability requirement, but still falls short of virtue. As James Cordell (2017) observes, for example, we could suppose that the society's director has deliberately selected its biased members to "achieve overall balance […] and a lively discussion" (p. 46, fn. 8). If we further supposed that the director reliably maintains this balance by intervening when membership changes threaten to skew the group's debates, the society's non-prejudiced character would be fairly stable. But unless the society were moved by non-prejudiced motives, Cordell contends, Fricker would not call the resulting quality a virtue. And indeed, Cordell asserts, the debate society's neutrality does not seem to be produced by neutral motives. Rather, it is a byproduct of prejudices that cancel each other out. This suggests a second constraint on our account of collective virtue, namely the requirement that virtuous groups are *responsive to virtuous motives*. Call this the *normativity requirement*.

Although both the stability and normativity requirements are valid constraints on an account of collective virtue, we have to be careful that they do not rule out too much. In particular, it cannot be a consequence of the normativity requirement that a group's *members* have to be responsive to virtuous motives. After all, we are appraising the character of group agents as subjects in their own right. In order to explain divergence cases of the kind discussed in section three, group-level motives must at least in principle be capable of diverging from individual motives.

This point emerges more forcefully when we shift our focus from the relatively mundane example of a debate society to more complex, and to my mind more interesting,

examples involving incorporated groups. Corporations typically have complex leadership structures and rely on a large workforce to implement executive decisions. As French (1979, 1995) argues in his influential account of corporate agency, an important function of such corporate structures is to facilitate decision making at all levels of the organization. To do this, corporate structures specify decision-making mechanisms that determine the roles and responsibilities of the members of an organization (Ritchie, 2020). These structures thus establish how corporate goals and decisions are formulated and who is authorized to speak or act on behalf of the corporation. In most cases, this allows corporations to function as relatively autonomous agents, with mechanisms in place that determine its beliefs, goals, and values independently of the beliefs, goals, and values of its workforce.

In order to assess the motives of a corporation, then, we must assess its corporate structure to find out what drives the corporation at the collective level. But what drives an organization at the collective level often comes apart from what drives its workforce at the individual level. If we are interested in the motives of an organization like IKEA, for instance, finding out what motivates its individual warehouse workers will not get us very far. We can learn more by studying what motivates IKEA's executive decisions: its mission statement, for example, and the goals and values anchored in its charter of association. Further, if we find that the corporate decisions enacted by IKEA at the collective level result from virtuous motives, the fact that its workforce is less than virtuously motivated should not lead us to conclude that IKEA *the corporation*'s motives are flawed. For the same reason, I think we should be open to the possibility that Fricker's debate society counts as virtuous if it is not an "invisible hand" that produces the group's neutrality, but the hand of an authorized representative that implements the *group's* motive to neutrality by applying selection at the gate. The challenge, then, is to explain how groups can be responsive to virtuous motives *at the group level*.

From my critique of plural subject accounts emerges a third, and final, constraint on accounts of collective virtue. If we want to preserve the intuition that the possession of virtue is deeply praiseworthy, manifesting group virtues cannot be too easy. We have seen, for instance, that the evaluative standard usually associated with motivation-based virtues is not met by traits that are produced by single acts of will. Instead, the presence of these virtues usually indicates that someone has put sustained effort into cultivating virtuous dispositions. The level of praise attached to motivation-based virtues thus reflects the sense in which we are responsible for developing a virtuous motivation. The virtuous agent's motivation is praiseworthy not because they have displayed good moral judgment once, but because they

are responsible for entrenching their good moral judgment over the course of many interactions with their environment. I will call this last constraint the *responsibility requirement*. The responsibility requirement entails that an organization or group must be responsible for entrenching its virtuous character traits if these traits are to be praiseworthy in the same sense as individual motivation-based virtues.

To sum up, a motivation-based account of collective virtue should ideally meet three constraints. First, it should meet the *stability requirement*. Just like their individual counterparts, collective virtues should be relatively immune from sudden changes. Second, it should meet the *normativity requirement*. Any motivation-based account must explain how groups can be responsive to virtuous motives. And third, it should meet the *responsibility requirement*. To maintain the usual evaluative connotations of motivation-based virtues, an account of collective virtue must explain how groups can be responsible for entrenching a virtuous motivation over time. With relevant substitutions, roughly the same constraints apply to an account of collective vice. Collective vices may not only involve vicious motives, however, but also the culpable failure to implement virtuous motives. As I will now show, a functionalist account of collective virtue and vice can meet each of these constraints.

2.5.2 Functionalism about Group Agency

On a functionalist account of group agency, groups have agency to the extent that they are organized as systems that *function* as agents (List & Pettit, 2011). In keeping with functionalist theories in the study of individual cognition, this entails viewing agency and associated mental states not in terms of their internal constitution, but in terms of how they *operate*. Christian List and Philip Pettit (2011) develop this view using a standard belief-desire conception of agency. According to this model, the possession of agency requires (i) representational states (or *beliefs* that depict the environment), (ii) motivational states (or *desires* as to how the environment should be), and (iii) the capacity to act on these motivational states in line with these representational states (see also List, 2018). In human beings such mental states as beliefs and desires are implemented by neurons and firing synapses, but functionalists maintain that nothing hinges on this. If systems that are otherwise constituted can play the functional roles associated with these states, we can ascribe the functional equivalent of these states to these systems as well. As List and Pettit observe, for example, even a basic thermostat may qualify as an agent on this account: it represents the

temperature in a certain room, aims at (or desires) a certain target temperature, and has the capacity to adjust the room's temperature in order to realize the target temperature.

Corporations and other collective actors can realize these conditions for agency as well. To illustrate, let us return to the case of *Honest Tobacco*. As with most corporations, *Honest Tobacco* is driven by a desire to turn profit. This desire to be profitable *motivates Honest Tobacco*'s investment decisions and market behavior. But the organization does not pursue its profit motive blindly. Its investment decisions and market behavior are *informed* by certain beliefs *Honest Tobacco* has about the market, about the product it is selling, and so on. Further, *Honest Tobacco* has the capacity to *act* on these desires in line with these beliefs. To maximize its returns, the company will only sell those products it believes will succeed. Bracketing, for a moment, the question of how these corporate beliefs, desires, and capacities for action are instantiated, they play the same functional roles as individual beliefs, desires, and capacities for action. We have, then, a system that functionally counts as an agent: *Honest Tobacco* exhibits representational states, motivational states, and the capacity to act on these.

According to functionalists, group agency and associated mental states are just as real as individual agency and associated mental states. Hence, List and Pettit (2011) describe themselves as "realists" (p. 6) about group agency. While individual beliefs, desires, and other agential structures are concentrated in a biological brain, however, *group agency* is implemented by group members in boards, teams, and other configurations. These members transmit information not, as in a biological brain, by firing synapses, but analogously by way of emails, memos, and other forms of communication. And they process this information not along neural pathways, but according to formal or informal decision-making structures adopted by the group (see, e.g., French, 1979, 1995; List & Pettit, 2002, 2004). As we have seen, these decision-making structures establish the mechanisms by which the group's views, goals, and executive decisions are formulated. Once formulated, the group depends on its members to act as proxies and carry out its views and decisions (Ludwig, 2014).

Functionalists, in other words, preserve the intuition that group agency ultimately depends on the behavior of group members. This is important to emphasize, because critics may worry that talk of group mental states is "metaphysically spooky" (Fricker, 2010, p. 242), introducing "psychologically mysterious forces" (List & Pettit, 2011, p. 4). We can now see, however, that this is not so. Adopting a functionalist point of view does not entail ontological commitments other than those involved in the claim that a group of *individuals* can be organized such that they collectively manifest the functional equivalent of certain mental states and other agential features. The beliefs, desires, and intentions of *Honest Tobacco* are

ultimately a function of the activities carried out by its employees. Depending on its corporate structure, its beliefs about the market, for example, may be formulated by its market analysts, while its goals and values may be set by management.

Quite another sort of critic, however, may object that a functionalist account of group agency is not collective enough. Since functionalists maintain that group agency is implemented by individual persons, one may worry that functionalists fail to explain how groups could be autonomous agents *in their own right*. Instead, group agency simply appears to be an extension of individual agency. If it is not clear that groups can be agents in their own right, however, it is not clear that we need an account of group virtue and vice either. We could probably dispense with extant accounts of individual virtue and vice and suitably generalize these along summative lines.

In their defense of functionalism about group agency, List and Pettit (2011) convincingly show this objection rests on a mistake. It is true, they write, that group agency and associated mental states "supervene on the contributions of its members" (p. 66). While this means that there cannot be a difference in group-level features without there being a difference at the individual level, the relation between individual and group-level features is highly complex. This complexity is due in part to the fact that groups can implement a large variety of "aggregation functions" (p. 60) that generate group-level intentional attitudes on the basis of profiles of individual attitudes. The group's organizational structure could include decision-making mechanisms that generate group-level beliefs by way of simple majoritarian voting procedures, for instance, but it can also opt for aggregation functions that generate group-level attitudes in other ways – say, by nominating a dictatorial CEO. Though certain aggregation functions maintain a close link between individual attitudes and group-level results (e.g., ones that require unanimity), other functions allow for significant discrepancies. In limit cases, List and Pettit show, an organization can come to endorse views not held by *any* of its members.[38] This suggests that functionalist group agents can exhibit a surprising degree of autonomy. And indeed, since group attitudes often come apart from member attitudes in unpredictable ways, List and Pettit conclude that group agency does not straightforwardly reduce to individual agency.

---

[38] To substantiate this argument, List and Pettit (2011) draw on so-called "impossibility result[s]" (p. 42) in social choice theory. These results show that it is impossible to generate rational group attitudes *unless* these attitudes can come apart from member attitudes. See List and Pettit (2002, 2004) for more.

2.5.3 Modeling Collective Motives: The Normativity Requirement

Adopting a functionalist point of view allows us to conceive of groups as agents in their own right, but the question remains whether groups can function as *virtuous* or *vicious* agents. Answering this question requires determining whether groups can manifest the functional equivalent of a virtue or vice. Specifically, we want to find out if groups can instantiate *motivation-based* virtues or vices. As I argued earlier, on a motivation-based picture virtuous or vicious functioning is subject to three constraints: the *stability*, *normativity*, and *responsibility* requirements. Starting with the normativity requirement, I discuss each of these constraints in the subsections that follow.

In order to meet the normativity requirement, we must show that group agents can be moved by virtuous or vicious motives. This requires demonstrating that groups can be responsive to virtuous or vicious reasons. Fortunately, a functionalist account of group agency has the resources we need. To see this, recall that motivational states are a crucial component of functionalist theories of group agency. The functional role of these motivational states is twofold. First, they represent the environment as the agent would *like* it to be. Second, they motivate the agent to bring about this desired state of affairs. In other words, they give the agent a reason for action – a *motive* to act in a certain way. For *Honest Tobacco*, that motive is profit-maximization, but it is possible for a group to adopt all sorts of motives.

Having shown that group agents can exhibit motives, it is fairly straightforward to demonstrate that these motives can be virtuous or vicious. In functionalist terms, a group agent exhibits virtuous or vicious motives when it realizes a motivational state that depicts, and therefore aims at, certain virtuous or vicious ends. Since most theorists agree that having a virtuous motivation involves pursuing virtuous ends *for their own sake* (*cf*. Johnson King, 2020), these virtuous ends must not be pursued for ulterior motives, but rather because they are *intrinsically valuable* (Hursthouse, 1999; Zagzebski, 1996). Given that virtues typically embody differing values, the content of these motivational states differs from virtue to virtue. An *honest* motivation involves responding to such values as honesty and transparency, for instance, whereas a *generous* motivation involves adopting such values as charity and kindness. The exact mechanism by which a group realizes these motivational states depends on its organizational structure; specifically, on how the group formulates its goals and desires. As we have seen, corporations like *Honest Tobacco* typically set their goals and values at the

management level. Accordingly, an organization like *Honest Tobacco* may realize a virtuous motivational state if management has voted to embrace morally good values.

　　To complete the case that a functionalist account can meet the normativity requirement, we must now show that group agents can be *responsive* to virtuous or vicious motives. After all, an agent can *have* virtuous motives, but fail to be moved by them.[39] Indeed, collective vices often consist in a *culpable failure* to implement virtuous motives. In his study of the global financial crisis, for instance, Boudewijn de Bruin (2015) finds that the motives of financial services providers were not blameworthy. It is just that these service providers were often "incompetent" at implementing virtuously motivated due diligence protocols. Fricker (2020) has recently discussed a similar vice in her study of "inferential inertia." Focusing on allegations of sexual misconduct surrounding the BBC, she shows how the BBC had both evidence of this misconduct *and* a motive to address it, but failed to respond properly due to various institutional pressures. Chapter four of this dissertation discusses a similarly vicious state of affairs at the American plane manufacturer Boeing.

　　For each group that fails to act on its motives, however, there are others whose motivational states do lead to appropriate action. Indeed, there are many steps a group can take to ensure that its motives receive adequate uptake. After embracing such values as honesty and transparency, for instance, a corporation like *Honest Tobacco* can see to it that these values are integrated into business practice by drawing up new decision-making protocols and ethics codes. To the extent that the company's adopted values would subsequently come to serve as reference points in corporate decision making – standards by which e.g. new business proposals are pursued or cast aside – these values would then function as motives that drive the corporation's actions.

　　Although a corporation relies on its *members* to carry out its decisions, moreover, none of this requires that these members identify with the organization's motives too closely. What matters is simply how the group functions at the collective level. So long as we can say that the *organization* authorizes its members to perform such-and-such action for such-and-such virtuous reason, we can attribute the organization's conduct to a virtuous motivation. Suppose, for example, that a retailer like IKEA announces a new recycling program after embracing environmental values. If these values truly dispose IKEA to act for virtuous reasons, a functionalist analysis allows that this program is virtuously motivated even if the

---

[39] In the study of individual virtue, this phenomenon is often discussed in terms of *akrasia* – weakness of the will.

warehouse workers that carry it out (e.g., by recycling used cardboard) do not care about the environment at all.

In short, embracing functionalism enables us to explain how group agents can be response to virtuous or vicious motives. This is sufficient to establish that a functionalist account of collective virtue and vice meets the *normativity requirement*. But can a functionalist account explain the sense in which virtues or vices are *counterfactually stable*?

2.5.4 Modeling Counterfactually Stable Group Traits: The Stability Requirement

In order to meet the stability requirement, an account of collective virtue and vice must show that groups can manifest counterfactually stable group traits. Specifically, it must demonstrate that groups can be disposed to respond to virtuous or vicious motives across a wide range of circumstances and in the face of countervailing motives. As we saw in my critique of Fricker, for instance, *Honest Tobacco* would not count as virtuous if it ceased to respond to virtuous motives the moment its executives reverted to dishonest values.

Although I have argued that plural subject accounts fail to meet the stability requirement, Fricker contends that these accounts are, in fact, uniquely well-positioned to explain the stability of collective virtues and vices. This, she claims, is because Gilbert's plural subject theory is "distinctive in embedding a strong notion of commitment in the very mechanism of group agency" (Fricker, 2020, p. 94). Insofar as the members of a plural subject *commit* to a certain ethos, Fricker suggests, that ethos and its values will continue to guide member behavior as long as the joint commitment persists. But in theories of group agency that are "less demanding in the dimension of commitment" (p. 1328), Fricker (2013) claims, "nothing internal to [the mechanism of group agency] would fix the group's efforts in positive relation to a value with the requisite stability" (p. 1329).

While it is true that a functionalist account does not include commitments as constitutive elements of group agency, it has room for other mechanisms by which collectives can realize counterfactually stable values and motives. To ensure that its values continue to motivate the organization in the face of countervailing motives, an organization could enshrine these values in the organization's constitution, corporate charter, or corporate by-laws, for instance. In some countries, including the Netherlands, the articles of association that establish a corporation with legal personality must include corporate objectives that are legally binding (*Burgerlijk Wetboek Boek* 2, art. 177). Since executives open themselves to

litigation if they act contrary to these objectives (*Burgerlijk Wetboek Boek 2,* art. 9*)*, this provides a strong incentive for executives to stay in line.

Moreover, amending these documents is often costly, time-consuming, and requires the approval of stakeholders other than executives. This makes it significantly harder to repudiate the values anchored in these legal documents on a whim. Organizations could also appoint independent oversight boards, ombudspersons, or create boards of trustees tasked with protecting the organization's values. These bodies could ensure that the organization's value commitments are upheld by scrutinizing the organization's conduct, offering counsel, and issuing (binding) advisories. Further, an organization could encourage its members to promote its values by creating incentives in the form of promotion schemes and performance bonuses; it could also impose sanctions for non-compliance. *Honest Tobacco*'s executives may be significantly less likely to embrace dishonest motives, for instance, if doing so would cause them to miss out on their end-of-year bonus.

In fact, functionalism allows for *myriad* ways in which an organization can stabilize its adopted values and therefore its disposition to act from virtuous or vicious motives. The brief overview of steps an organization can take to insulate its values from sudden changes enumerated in the previous paragraphs is far from exhaustive. On a functionalist analysis, what matters is simply that an organization *functions* in such a way that it is robustly disposed to act from virtuous or vicious motives. *How* the organization implements this disposition – i.e., by which organizational structures this disposition is realized – is not important.

On the face of it, then, a functionalist account of collective virtue and vice meets the *stability requirement* as well. But a familiar objection looms in the background. Fricker (2013) addresses this objection when she assesses the importance of such stabilizing mechanisms as incentives in the form of promotion schemes and performance bonuses. While these can work well "in some areas [of good institutional performance]," Fricker worries they fall short in areas where good performance hinges on things "being done *in the right spirit*" (p. 1326, emphasis mine); specifically, in areas where *motives* are important. Presumably, the thought here is that a group member who promotes the group's virtuous ends only to achieve e.g. a promotion or other performance bonus does not act as an agent of virtue, but rather as an agent that is selfishly motivated. In that case, how can the actions the agent performs on behalf of the group be attributed to virtuous motives?

We can neutralize this objection by once more attending carefully to the relationship between a group agent and its members. In settling the question of whether a group agent exhibits counterfactually stable virtuous traits, it does not matter if its *members* function as

agents of virtue, but it matters whether the *group as a whole* does. A group can realize a robustly stable motivational state, however, *regardless* of the motives of its members. Indeed, so long as a group can reliably get its members to implement the group's virtuous motives (by carrying out its decisions, and so on), that group could theoretically realize counterfactually stable virtuous motivational states even if it were entirely composed of vicious members.[40] I think it is plausible, in fact, that the most successful groups exploit this feature by ensuring that its vicious members are placed in roles where their character flaws can be beneficial to the organization. Though *overconfidence* is a canonical epistemic vice, for instance, there is some empirical evidence that it can be a valuable trait in CEOs, contributing to their company's success and efficiency (de Bruin, 2015, pp. 118-122; Hirshleifer, Low, & Teoh, 2012).

To sum up, a functionalist account of collective virtue and vice has little difficulty modeling counterfactually stable group traits. Accordingly, it meets the *stability requirement* as well. Taken together, meeting the normativity and stability requirements entails that a functionalist account can explain how groups can be robustly disposed to act from virtuous or vicious motives. Since plural subject accounts ultimately fail to meet the stability requirement, this is sufficient to show that a functionalist account is preferable. In the next subsection, I assess whether a functionalist account can meet the third and final constraint. It would be highly desirable if our account could explain the usual evaluative implications of collective virtue and vice attributions.

2.5.5 A Role for Organizational Routines: The Responsibility Requirement

The third, and final, constraint on our account of collective virtue and vice takes the form of a subjunctive conditional: *if* we wish to preserve the intuition that virtues are deeply praiseworthy and vices deeply blameworthy, it should explain the sense in which organizations can be responsible for cultivating these traits over time. At any rate, it should not make cultivating these traits too easy – a matter of a single act of commitment. I formulate this constraint as a subjective conditional because I take it that one may deny its antecedent. Talk about motivation-based virtues and vices is interesting for its explanatory powers quite independently from its usual evaluative connotations. As Candace Upton (2009a) points out, "trait attribution is useful not only for moral appraisal, but it also serves as a source of

---

[40] Hence, functionalism can accommodate the divergence cases discussed in section two.

information, explanation, and prediction" (p. 185) insofar as we expect virtuous or vicious individuals to engage in characteristic types of behavior.

Nevertheless, since it is highly desirable for an account of virtue and vice to explain the evaluative importance usually attached to virtues and vices, I sketch a way for a functionalist account to meet this evaluative challenge in what follows. As it stands, we have identified motivation-based virtues and vices with the functional state of an organization. But this leaves open the question of whether that functional state is easy to realize or difficult. It says nothing, for instance, about the etiology of virtuous or vicious functional states, which is what often justifies the deep praise or blame attached to ascriptions of individual virtues and vices.

As a first step, however, observe that realizing a functional organization that corresponds to a collective virtue or vice is not typically an easy feat – certainly not as easy as jointly committing to virtuous or vicious motives. To ensure that an organization is robustly functionally responsive to virtuous or vicious motives, that organization typically needs to enact a strong institutional infrastructure that guides the organization's conduct. This involves creating decision-making mechanisms that are responsive to virtuous or vicious values and reliably translate into virtuous or vicious practice. Codes of conduct, decision-making protocols, and other policies must be drafted, and the organization's members must be convinced to use these in practice. As the extensive management literature can attest to, getting this institutional framework right is remarkably challenging; building an organizational structure that works is often a continuous process of trial and error.

In the field of organizational studies, the sense in which institutional well-functioning is an ongoing process is reflected in the study of "organizational routines" (Becker, 2008; Nelson & Winter, 1985). As Marcus Becker (2004) observes in his widely influential review, these routines are the organizational counterpart of individual "habits" (p. 645): rule-bound, recurrent patterns of activity at the organizational level. Examples of organizational routines comprise anything from certain ways of conducting sales to particular patterns of behavior on the manufacturing floor. Though an important function of these routines is to bring stability to organizations, they are subject to gradual changes over time (Feldman, 2000; Feldman & Pentland, 2003; Pentland, Hærem, & Hillison, 2011). At each iteration, "[p]eople have a choice between whether to [reproduce an organizational routine]," Becker (2004) writes, "or whether to amend the routine" (p. 648). Organizational routines are not reproduced blindly, then, they are continually adapted in response to situational needs and based on previous outcomes. As a consequence, organizational routines are important vehicles for

"organizational learning" (Levitt & March, 1988; Tsoukas & Chia, 2002) – mechanisms by which an organization can strive to improve its functioning over time.

My contention is that organizational routines can play the same role in explaining why collective virtues and vices are deeply praise- or blameworthy as habits in the case of individual virtues and vices. Just as individual virtue or vice is often the result of long processes of habituation and learning, that is, a collectively virtuous or vicious functional state often reflects the presence of virtuous or vicious organizational routines; routines that explain *why* an organization is disposed to virtuous or vicious behavior. And like habits, these routines have a history – after all, "one would be hard pressed to call something happening only once a routine" (Becker, 2004, p. 646). They typically start out as relatively vulnerable patterns of behavior that become more stable, and indeed *entrenched*, as they are reproduced and their norms are reified. Crucially, moreover, this history represents a learning curve: throughout its life, an organizational routine undergoes numerous changes in response to challenges from the environment. Sometimes these changes are for the better, as when an organization finally settles on non-prejudiced hiring routines after a long history of sexism or racism. At other times, an organization learns the wrong lessons, as when a history of getting away with fraud gradually cements that organization's fraudulent disposition.

When a virtuous or vicious functional state coincides with certain organizational routines, then, that functional state is evidence of a *history* of good or bad judgment. The group *could have* learned the wrong lessons in the course of reproducing its routines, but it in fact learned the right ones (or vice versa) – a significant achievement (albeit a negative one in the case of vice). This is precisely the dynamic that justifies the deep praise or blame attached to individual virtues and vices as well. The acquisition of virtue is typically taken to be "a very gradual process full of numerous setbacks" (Miller, 2003, p. 378), requiring "lifelong effort, in some cases getting things wrong and then learning from mistakes" (Kupperman, 2001, p. 243). Hence, in Zagzebski's (1996) words, "the particular kind of praiseworthiness that applies to a virtue […] reflects the fact that the virtuous person might have been vicious [had they developed the *wrong* habits] instead" (p. 105). Or as Hursthouse (1999) puts it, to attribute a virtue to someone is to go "far beyond the moment of action" (p. 123) and extend praise also to the kind of person someone has fashioned themselves into over time. I submit that virtuous or vicious functional states that correspond to the presence of virtuous or vicious organizational routines are deeply praise- or blameworthy in the same regard. Cultivating these routines requires time, effort, and learning. In other words, cultivating these routines is far from easy.

Furthermore, the idea that organizational routines often play a part in virtuous or vicious organizational functioning has considerable empirical purchase. One common theme in the business ethics literature, for instance, is that corporations often *gradually* collapse into vicious patterns of behavior. The most (in)famous example of such a collapse involves the American energy trader Enron. In their classic study, Ronald R. Sims and Johannes Brinkmann (2003) characterize Enron's descent into corporate vice as a process of gradual "ethical erosion" (p. 245). Though Enron started out as a paragon of corporate virtue, with thorough ethics codes, an extensive corporate citizenship program, sound incentives, and organizational routines that embodied these virtues, Enron's employees gradually shifted away from these to a "culture of greed" (p. 241). Sims and Brinkmann describe this shift in detail: with every ethical norm broken in the pursuit of profit, it became easier for Enron's employees to break other norms, until the organization finally settled on routines that embodied a total disregard for ethical norms and values. Once Enron's ethical norms had withered away completely, what remained was a company with a deeply entrenched disposition to engage in fraudulent behavior.

In light of the above, I contend that a functionalist account of collective virtue and vice can meet the *responsibility requirement* as well. Specifically, by invoking the concept of organizational routines a functionalist account can explain why we hold collectives deeply responsible, and therefore view them as deeply praise- or blameworthy, for cultivating a virtuous or vicious functional state. On the picture that emerges, a group agent counts as virtuous or vicious in a deeply praise- or blameworthy sense if it manifests the functional equivalent of a virtue or vice *as a consequence* of cultivating virtuous or vicious organizational routines. By contrast, if a group manifests a virtuous or vicious functional state without such causal history, it will be robustly disposed to act from virtuous or vicious motives, but the quality of praise or blame attached to that disposition is severely diminished. Indeed, the dispositions of such a group will be closer to the dispositions of someone whose character has been transformed by Nozick's transformation machine than those of an actual virtuous person.

2.6 Enriching the Account: Group Emotions

The preceding sections show that a functionalist account of collective virtue and vice satisfies three important constraints; it can model virtuous or vicious collective motives, stable

dispositions, and explain the evaluative importance attached to virtuous or vicious functional states. Since plural subject accounts do not meet each of these constraints, this establishes a functionalist account as the preferable alternative. But there is another advantage to adopting a functionalist account – its remarkable versatility. Adopting a functionalist point of view allows us to model not just collective motives and dispositions, but the collective equivalent of *any* psychological structure whose functional role can be realized within groups. As I will now show, this opens up promising avenues for future research.

An immediate payoff of embracing functionalism is that we can enrich the study of collective virtues and vices by incorporating *collective emotions*. Though existing research into group virtues and vices rather narrowly focuses on the importance of motives, traditional virtue theorists examine a far wider range of themes that includes the virtuous (or vicious) person's emotional life. Characteristic of the virtuous person, for instance, is that she exhibits situationally appropriate emotions and reactive attitudes (Brady, 2018; Hursthouse, 1999; Price, 2011; Sreenivasan, 2020; Swanton, 2003). Zagzebski (1996) illustrates this point with respect to the virtuous person's response to so-called "tragic dilemmas" in which she is forced to sacrifice one of her values. If no matter what she does she violates one of her values, the virtuous person "will feel guilt and remorse" (p. 241), writes Zagzebski, rather than such inappropriate emotions as relief or complacency.[41] In the domain of vice, Hursthouse (1999) illustrates the importance of our emotions through a careful study of the vice of racism, which expresses itself not only in such emotions as "hatred and contempt", but also in "fear, anger, reserve, suspicion" (p. 114) and so on. If we want our account of collective virtue and vice to mirror the richness of individual accounts, then, it would be nice if we could say something about group emotions as well.

Though there is a vibrant literature on group emotions (Konzelmann Ziv & Schmid, 2014; von Scheve & Ismer, 2013), research connecting group emotions to group virtues and vices is virtually non-existent. This lacuna possibly reflects the fact that much of the literature on group emotions has an individualist bent, studying not how *group agents themselves* can manifest emotions, but the emotions *individuals* experience in, or on behalf of, groups (Smith, Seger, & Mackie, 2007). Paradigmatic examples of such emotions include the pride a saxophone player feels on behalf of her orchestra's performance, for example, or the contagious joy experienced in a celebrating crowd. A similarly individualist approach to the study of group emotions is implicit in the work of many social ontologists. Perhaps fearing

---

[41] See also Swanton (2003, p. 31).

that any account on which group agents can *feel* things is implausible (List, 2018) or counterintuitive (Knobe & Prinz, 2008), for instance, List and Pettit (2011) flat-out assert that "group agents […] lack the […] emotions of human individuals" (p. 176).

More recently, however, a number of functionalists have argued that group agents can be in genuine emotional states as subjects in their own right (Björnsson & Hess, 2017; Huebner, 2011; Schwitzgebel, 2015). In true functionalist spirit, the success of these arguments hinges on viewing emotions not in terms of their internal constitution or accompanying phenomenology, but in terms of how they operate – their *functional role*. Bryce Huebner (2011) illustrates this approach by describing a ship crew that collectively displays the emotion of fear. Very roughly, Huebner maintains, the functional role of fear is twofold: fear (a) *represents* a certain danger, and (b) *facilitates* "distinct action-tendencies" (p. 97) in response to perceived threats, for example by initiating bodily changes that allow us to escape. Groups, Huebner claims, can be organized to function in the same way. Upon learning that a ship's engines are failing, for instance, the ship's crew can come to believe the ship is in danger of running aground (in line with [a]). This representational state will in turn facilitate all sorts of action-tendencies that prepare the crew to deal with that perceived danger (in line with [b]): the ship's speed will be adjusted downwards, its trajectory will be monitored more closely, engineering crews will be dispatched, et cetera. The ship's crew, in other words, can be organized as a system that occupies *the same functional role* as the individual emotion of fear. And crucially, no crew member need be *individually* afraid. They may simply be going through protocols, failing to represent the signals the crew is collectively responding to *as* dangerous. On a functionalist analysis, this is sufficient to establish that groups, too, can be in emotional states.

If groups can manifest the functional equivalent of emotional states, it seems only natural to suppose that they can manifest these emotional states appropriately or inappropriately. While the state of fear described by Huebner can help the ship's crew deal effectively with danger, for example, we can also imagine a scenario in which the ship's crew experiences fear disproportionally. If the ship's crew reallocates *too many* resources in response to a perceived engine failure (e.g., by directing all crew to the engine room or issuing a premature order to evacuate), for instance, the ship's operations may be severely impeded and so hamper an effective response. Similarly, while it is often appropriate for groups like corporations to express moral emotions such as guilt (Björnsson & Hess, 2017), corporations can also do so inappropriately, as when guilt over a minor transgression is blown out of proportion and leads a corporation to shutter its doors, leaving employees and

customers that depend on it out in the cold. This suggests that just as the virtuous individual, a virtuous organization must learn to manifest proportionate and situationally appropriate emotions. It is at least *prima facie* plausible, then, that collective virtues have a role to play in regulating the emotional life of their possessor – perhaps as "correctives" (Foot, 1978, p. 8) that steer organizations away from excessive emotional displays.

As this brief discussion indicates, embracing functionalism gives us the resources to construct a rich account of collective virtue and vice that includes a prominent role for collective emotions. As such, it reveals an interesting range of new topics to be explored. Which collective emotions are expressive of what collective virtues or vices? How do virtuous functional states regulate a group's emotional life? And conversely, how do vicious functional states *disturb* a group's emotional life? But these questions are just the tip of the iceberg. Beyond motives and emotions, virtue theorists study attitudes of love, respect, and creativity (Swanton, 2003), modes of perception (McDowell, 1979), and distinctive dispositions to notice and expect in certain characteristic ways (Hursthouse, 1999), to name just a few common themes. To the extent that we can model these attitudes, modes, and dispositions along functionalist lines, a functionalist account of collective virtue and vice promises an even richer conception of collective virtues and vices – fertile ground, then, for future research.

2.7 Conclusion

One of the main attractions of motivation-based (or responsibilist) virtue theory is the descriptive richness of virtue and vice attributions. Calling a person *courageous* or *greedy* goes beyond assessing the value of that person's actions by evaluating their character, the kind of person they have fashioned themselves into. This entails attributing to that person a range of highly specific motives, dispositions, and values, qualities that are often thought to be deeply praise- or blameworthy.

Responding to recent developments in social ontology, a number of virtue theorists have attempted to construct a similarly rich account of *collective* virtue and vice. Their goal has been to explain how we can make sense of the notion of group or institutional character, where group character is thought to embody a similarly praise- or blameworthy range of motives, dispositions, and values. I have argued that any such account that wishes to model motivation-based virtues or vices must meet three conditions: the stability, normativity, and

responsibility requirements. After critically examining existing accounts, I have argued that a *functionalist* account of collective virtue and vice is preferable. On the view I develop, group virtues or vices arise when organizations or other collective actors function *as* virtuous or vicious. In the case of motivation-based virtues or vices, this requires that they are functionally organized so as to implement a counterfactually stable disposition to act from virtuous or vicious motives (or, in the case of some vices, that they culpably *fail* to implement a virtuous motivational state). A functionalist account can not only explain how groups can instantiate deeply praise- or blameworthy, counterfactually stable dispositions to act from virtuous or vicious motives, but also promises to enrich the study of collective virtue and vice by incorporating the collective equivalent of other mental states such as group emotions.

Alas, a functionalist account of collective virtue and vice is beset by problems of its own. While functionalist accounts of group agency draw parallels with *individual* psychological structures, a number of empirically-minded philosophers have recently argued that social psychological research reveals there to be *no such things* as individual character traits. If their argument is correct, it not only calls into question the very foundations of the virtue theoretical project, but also suggests that we should avoid drawing *too* close a parallel between individual and collective virtues and vices. In the next chapter, I examine whether and how we can accommodate empirical challenges to the virtue theoretical project and what their implications are for a functionalist account of group character.

Chapter 3 – Expanding the Situationist Challenge to Collectives

3.1 Introduction

Suppose you find yourself walking down a busy street and encounter a toddler, lying face down in the middle of the road. What would you do? If you are like most people, you will answer this question by saying that of course, you would *help* the toddler. As a matter of fact, however, there is a good chance you would remain passive, at least initially. Such was the fate of two-year-old Yueyue, ignored by scores of passersby after she was struck by a car in the Chinese city of Foshan.[42]

If the story of Yueyue surprises you, consider that the so-called "bystander effect" is one of the strongest findings in all of social psychology (Fischer et al., 2011). Time and time again, experimenters have shown that the chance someone will help a stranger in need decreases in proportion to the number of onlookers present. In so doing, their research fits squarely within the tradition of *situationism*, a school of thought that highlights the impact of seemingly trivial situational influences on our behavior.

According to situationists, the impact of situational influences on our behavior casts doubt on the existence of robust character traits (Ross & Nisbett, 2011). As critics like John Doris (1998, 2002) and Gilbert Harman (1999, 2000) point out, these findings thus apparently call into question the empirical basis of contemporary virtue ethics. The crux of Doris's and Harman's objection lies in the fact that we simply do not observe the cross-situational behavioral regularities the widespread existence of robust traits like virtues and vices would predict. Indeed, Harman emphasizes, we often erroneously attribute behavioral outcomes to virtues and vices – a phenomenon social psychologists call "the fundamental attribution error" (Ross, 1977, p. 183).

Unsurprisingly, these so-called *situationist critiques* of virtue ethics have generated much debate among virtue theorists (Appiah, 2008; Upton, 2009b). Though most of this

---

[42] For more on the tragic story of Yueyue, see Osnos (2011).

debate has a strikingly individualist bent, collective phenomena often play an important role in the situational dynamics under discussion. The bystander effect illustrates this nicely: it is a *crowd* of bystanders that can all but reduce our compassion to impotence. Accordingly, it would seem that virtue theorists stand much to gain from considering the situationist critiques in relation to questions of groups and group agency.

This chapter takes up that challenge by critically examining the role of collectives in the situationist literature. Two questions take center stage. First, what do situationist findings tell us about the impact of groups on the conduct of individuals? And second, what do these findings imply for the notion of group character, including the idea that groups can be virtuous or vicious as subjects in their own right? The result of extending the situationist challenge in these ways is a situationist program that consists of three levels. The first studies the impact of situational influences on individuals in isolation; the second studies the impact of situational influences on individuals as they combine in groups; and the third studies the impact of situational influences on groups as such.

As we will see, evaluating the situationist challenge from the perspective of groups and group agency pays off in two major ways. The first reward manifests itself on the level at which individuals interact with each other in groups. Taking situationist findings seriously, I argue, reveals that groups often facilitate the exercise of *local* (i.e., situationally-dependent) virtues and vices in individual group members. We can use this to our advantage when we consider questions of organizational and institutional design. The second pay-off manifests itself at the collective level and concerns the concept of group character. Since group agents do not face the psychological limitations of individuals, they are subject to different situational influences. Expanding the situationist challenge to group agents therefore reveals the need for a new research program that studies the behavioral variability of collectives under various situational pressures. Once that program is underway, it promises to deliver a richer picture of the sorts of dynamics that can destabilize virtuous or vicious dispositions at the collective level.

My strategy is as follows. The next two sections concern situationism at the *individual* level. In section two, I introduce the challenge posed by situationism in greater detail. As we will see in section in three, virtue theorists have responded to this challenge in various ways. One particularly promising response centers on the importance of *local* character traits. Section four discusses the implications of this strategy for the role groups play in mediating virtuous or vicious behavior, taking us to the level at which individuals interface with group

dynamics. Finally, in section five, I expand the situationist challenge to collectives *as such*. Section six concludes.

3.2 Situationist Critiques of Virtue Theory

Although virtue theory is a highly diverse field, most practitioners agree that virtues and vices are entrenched character dispositions to act or cognize in a particular way (Athanassoulis, 2012; Broadie, 1991; Hursthouse, 1999; Slote, 2001; Swanton, 2003; Zagzebski, 1996).[43] These dispositions are usually given a standard counterfactual analysis: if she *were* in a situation that required it, the virtuous person *would* display virtuous behavior. On a default understanding, then, we should expect a virtuous person to act in characteristic ways across a range of virtue-eliciting situations (Hursthouse, 1999, ch. 2), including ones she has never before encountered (Nussbaum, 1990). Someone with the virtue of honesty, for example, will be honest *no matter the circumstances* that require it; she will not deceive her boss to get a promotion, nor will she cheat on a math test.

As a number of ethical theorists have remarked, this standard virtue theoretical account appears to have empirical commitments that can be tested (Alfano, 2013; Doris, 1998, 2002; Flanagan, 1993; Harman, 1999, 2000). When we ascribe a virtuous or vicious trait to someone, we implicitly make certain predictions about that person's behavior in trait-relevant conditions. We predict that the compassionate person will be kind to strangers *as well as* friends in need, and that she will be kind in good *or* bad conditions. Such predictions can be put to the test by examining whether virtuous persons *in fact* display trait-relevant behavior across different situations – a phenomenon social psychologists call *cross-situational consistency* (Mischel, 1968, 2009). If virtues and vices are as widespread as our ordinary usage of virtue and vice terms would seem to suggest, moreover, we should expect to observe many such behavioral regularities in society at large.

According to situationist critics, however, social psychological findings overwhelmingly cast doubt on the empirical commitments of virtue ethics (Merritt, Doris, & Harman, 2010; Upton, 2009b) and virtue epistemology (Alfano, 2012; Olin & Doris, 2014). The most influential critiques are due to Doris (1998, 2002) and Harman (1999, 2000, 2009),

---

[43] Of course, there are notable exceptions. Judith Thomson (1997) argues that virtues and vices are properties of actions, not character; and in the domain of virtue epistemology, reliabilists argue that virtues and vices comprise *all* features that produce good or bad epistemic effects (Greco, 2010; Sosa, 2007). This does not render reliabilism immune from the situationist challenge (Alfano, 2014).

whose objections to virtue theory continue to shape the situationist debate to this day.[44] Though their critiques are characterized by subtle distinctions, both Doris and Harman argue experimental research fails to show that human subjects typically display cross-situational consistency to the degree required by the possession of virtue.[45] In fact, they claim, experimental research suggests most human subjects lack broad-based character dispositions altogether. While we are apt to attribute behavioral outcomes to underlying character traits, research within the situationist paradigm suggests that most of our behavior is a function of – often trivial – situational influences (Ross & Nisbett, 2011). Hence, we frequently commit the "fundamental attribution error" (Ross, 1977, p. 183) of attributing to character what is in fact the result of situational factors.

To substantiate their claims, Doris and Harman draw on a rich body of experimental research that studies the impact of situational influences on our behavior. One classic study that figures prominently in the work of Doris, Harman, and situationist critics that followed them involves a series of experiments conducted by Stanley Milgram (1974). Milgram's experiments, which were designed to test the effect of authority on our willingness to follow orders, famously showed that a majority of human beings appear willing to inflict significant harm on others when instructed to do so by someone in a position of authority. Since we would expect someone with a compassionate character to resist these instructions, situationists take this result to show that most human beings lack a compassionate character. But it is equally unlikely, situationists claim, that Milgram "stumbled onto an aberrant pocket of sadists" (Doris, 2002, p. 42). To explain Milgram's results, situationists therefore do not point to the character of the study participants involved, but rather to the impact of situational influences endemic to Milgram's experimental setting. These influences include the way instructions were communicated to study participants, the authoritative status of the experimenter communicating the instructions, and the manner in which study participants received feedback on their actions (Ross & Nisbett, 2011).

Another classic study that is often cited in support of situationist critiques is John Darley and Daniel Batson's Princeton Theological Seminary study (Darley & Batson, 1973). In Darley and Batson's experiment, study participants – all seminary students – were

---

[44] Though most influential, Doris and Harman are not the first philosophers to discuss situationist research in the context of virtue theory. Earlier work includes Flanagan (1993).

[45] Doris (1998) additionally objects to a second feature of (Aristotelian) virtue theory, namely the thesis that virtuous persons display *evaluative consistency* (which entails that virtuous traits are correlated with each other, such that an honest person is also likely to be e.g. generous and courageous). Since this objection does not figure prominently in the situationist debate, I leave it aside here.

instructed to deliver a talk on the other side of campus and told they were in varying degrees of hurry. On the way to the lecture hall, the experimenters had arranged for the study participants to encounter a clearly distressed person slouched on the floor. What Darley and Batson wanted to find out is whether the study participants would help this person, who was in fact a confederate involved in the study. Somewhat surprisingly, Darley and Batson found that helping behavior was correlated strongly with the degree of hurry the study participants were in, but not with other variables, such as their religious background or moral beliefs. Study participants who were *not* in a hurry overwhelmingly helped the distressed person, while participants who were in a rush overwhelmingly did not. This result is all the more poignant given that the study participants were to deliver a sermon on the parable of the Good Samaritan, which teaches the importance of kindness to strangers. Again, situationists conclude that situational factors – in this case, *being in a hurry* – provide a better explanation of this finding than character-based factors. After all, we would expect someone with a helpful or compassionate character to help a distressed person regardless of how much of a hurry they are in.

While the experiments conducted by Milgram, Darley, and Batson are among the most influential situationist experiments, there is a wide array of studies that report similar findings. Research on mood effects, for instance, suggests that the mood one is in drastically affects the degree to which one is likely to help others (Isen & Levin, 1972). The behavioral impact of mood changes is significant even when these changes are related to such trivial factors as pleasant smells (Baron & Thomley, 1994; Liljenquist, Zhong, & Galinsky, 2010), background music (Alpert & Alpert, 1989; Fried & Berkowitz, 1979), and being in good or bad weather conditions (Cunningham, 1979). More recently, psychologists Dan Ariely (2008) and Daniel Kahneman (2011) have drawn attention to a wide range of other situational factors liable to influence our behavior and cognitive processes. To foreshadow my later discussion, a significant number of situational influences involve group dynamics – from group roles (e.g., Zimbardo, 1973) to group conformity (Asch, 1955) and the bystander effect (Fischer et al., 2011). Far from indicating that we are capable of cross-situationally consistent behavior, situationists maintain, these findings suggest our behavior is highly variable from one situation to another. Accordingly, situationist findings are difficult to square with the notion that our behavior is a function of underlying character traits.

If the existence of robust dispositional character traits were as widespread as we tend to believe, we should not expect to see the widespread correlation of behavioral outcomes with situational influences these studies reveal. But why does this spell trouble for virtue

theorists? Harman (1999) takes these studies to show that "there is no empirical basis for the existence of character traits" (p. 316), and hence that virtue theory is empirically unfounded. While virtuous character traits *should* robustly dispose one to respond to ethically salient considerations, situationist studies of the kind just cited demonstrate that virtually everyone's behavior is apt to respond to such ethically insignificant situational influences as one's mood or degree of hurry. Moreover, Harman (2000) argues, the widespread absence of character traits renders virtue theory all but impotent – how can virtue theory give us any moral guidance if attaining virtuous character traits is impossible for most, if not all, ordinary human beings? Harman concludes that we are better off abandoning virtue theory.

Though equally critical of virtue theory in its present form, Doris (1998, 2002, 2009) takes a slightly different tack. Doris (1998) agrees that situationist research exposes virtue theory to "damaging empirical criticism" (p. 520) insofar as this research suggests broad-based character traits do not exist. But rather than suggesting we abandon virtue theory altogether, Doris leaves open the possibility of *rehabilitating* virtue theory by substituting its notion of character with one that is empirically adequate. In particular, Doris claims that social psychologists have found ample evidence for the existence of "local" (p. 507) character traits – cognitive or behavioral dispositions that produce consistent behavior in a narrow domain of relevantly similar situations. Hence, our overall character is highly "fragmented" (p. 508), producing consistent behavior in some, but not other situations. As we shall see shortly, a number of virtue theorists have used this notion of local character traits in attempts to save virtue ethics and epistemology from the situationists' most damaging objections. Their arguments have important implications for the role of groups and group agents in facilitating the exercise of virtues and vices.

3.3 Virtue Theory after Situationism

Since situationists attack the very foundations of the virtue theoretical project, their critiques have spawned rich debate among virtue theorists (Alzola, 2008; Appiah, 2008; Upton, 2009b). Most defenses of virtue theory challenge situationist critics on *conceptual* grounds (Annas, 2003, 2005; Athanassoulis, 2012; Bates, Snow, Bates, & Kleingeld, 2018; Kamtekar, 2004; Kupperman, 2001; Sabini & Silver, 2005; Sreenivasan, 2002). The common thread in these conceptual defenses is the idea that situationist critics employ an overly narrow

conception of virtue and thus miss their intended target.[46] Moreover, advocates of a conceptual defense maintain that situationist results do not undermine a more accurate understanding of virtue theory.

Those who defend virtue theory on conceptual grounds can adopt any number of strategies. For instance, a number of theorists argue that situationist experiments show something virtue theorists can readily admit: that the possession of virtue is exceedingly rare, and that most ordinary humans fail to act virtuously (Athanassoulis, 2000; Kamtekar, 2004; Miller, 2003; Wielenberg, 2006). Others claim that Doris and Harman work with an overly demanding conception of virtue, arguing that being virtuous does not require that one acts virtuously on *every* virtue-eliciting occasion (Kupperman, 2001; Miller, 2003).[47] Additionally, some defenders of virtue theory point out that its focus on outwardly observable behavior prevents situationist research from addressing the *internal* dimension of traditional virtues, which also encompasses characteristic motives, emotions, and beliefs (Annas, 2005; Athanassoulis, 2012; Sreenivasan, 2002). This leaves open the possibility that the participants in situationist studies were in fact moved by virtuous motives, conceived of their behavior as virtuous, or did not perceive the experimental setting as one that required the exercise of virtue.

Given the vibrancy of the situationist debate, this brief overview of conceptual defenses is far from exhaustive. It is easier to survey a second, more recent, strategy critics of the situationist challenge have adopted. A small but growing number of theorists contest situationist critiques on *empirical* or *methodological* grounds (Alzola, 2008, 2012; Cohen & Morse, 2014; Prinz, 2009; Snow, 2010; Sreenivasan, 2002; West, 2018). Miguel Alzola (2012) for instance, objects that while only iterated trials can detect cross-situational consistency, most situationist research concerns "one-shot" (p. 97) studies that assess the behavior of individuals on a single occasion. These single observations, Alzola claims, are an insufficient basis for making inferences about the characters of the participants involved. Others have challenged the validity of particular situationist experiments, some of which – such as the famous Zimbardo (1973) prison experiment – have recently become embroiled in

---

[46] Bates et al. (2018) charge that situationists work with an overly narrow conception of *vice*, arguing that situationist results are consistent with the common existence of "vices that do not involve other-directed malevolence" (p. 524).

[47] Moreover, as e.g. Alzola (2008) and Solomon (2003) emphasize, the kinds of situations in which a virtuous person may err would seem to be precisely the extreme and unfamiliar situations presented in such situationist experiments as the Milgram study. See also (Upton, 2009a): "The broad range of situations across which a virtuous person must behave virtuously should be normal" (p. 179).

social psychology's replication crisis (Blum, 2018). Meanwhile, yet others have turned to psychological evidence that appears to vindicate the existence of broad-based character traits or equivalent personality structures (Prinz, 2009; Russell, 2009; Snow, 2010; West, 2018).[48]

Although it will be interesting to see what results from these conceptual and empirical defenses – it is certainly too early to conclude that we ought to abandon virtue theory altogether – the remainder of this chapter focuses on a third strategy virtue theorists have pursued in response to the situationist critiques, one that potentially sheds light on the role of collectives in facilitating virtues or vices in individual group members. This strategy involves combining situationist research and conceptual analysis to develop an account of virtue and vice that accommodates the psychological complexity revealed by situationist experiments (Doris, 1998, 2002; Miller, 2003; Upton, 2005, 2009a; Vranas, 2005). In contrast to those who defend traditional virtue ethics on *empirical grounds* (Russell, 2009; Snow, 2010; West, 2018), proponents of this strategy accept that situationist research shows our characters to be highly fragmented; in contrast to those who defend traditional virtue ethics on *conceptual grounds*, proponents of this strategy argue that the fragmented nature of our characters encourages us to modify our conception of virtue. Call the resulting approach a *hybrid response* to situationism.

Inviting virtue theorists to substitute their notion of character with one that is empirically adequate, the earliest advocate of a hybrid response to situationism is Doris (1998, 2002) himself. According to Doris, situationist research calls into question the existence of global, but not "local" (Doris, 1998, p. 508) character traits. While the results of situationist experiments suggest human behavior lacks cross-situational consistency, Doris argues, they also reveal a surprising degree of behavioral consistency in situations that are relevantly similar. The Princeton Theological Seminary study, for instance, suggests that most human beings lack the cross-situationally consistent trait of *compassion*, but also suggests that most human beings are compassionate when they are not in a hurry. Similarly, research on mood effects demonstrates that most people are consistently disposed to helpful behavior when they are in a good mood, but not when they are in a bad mood. According to Doris, these behavioral regularities support the existence of such local traits as *compassionate-when-not-in-a-hurry* or *helpful-in-a-good-mood*. To the extent that someone is disposed to virtuous

---

[48] Though Jesse Prinz (2009) challenges Doris and Harman on their empirical claims, he does not believe this helps the virtue ethicist much. While Prinz argues that broad-based character traits exist, he denies that they can play the normative role required by traditional virtue ethics.

or vicious behavior in *certain situations*, then, we can ascribe to that person a *local* virtue or vice that is indexed to the kind of situations in which that trait manifests itself.

Since the existence of local character traits enjoys widespread empirical support (Ross & Nisbett, 2011; Shoda, Mischel, & Wright, 1994; Wright & Mischel, 1987), it is not seriously disputed even by those who object to situationism on empirical grounds. More controversial is the conceptual question of whether local character traits are an adequate substitute of global traits in theories of virtue. Doris (1998, 2002) is optimistic: though not fully equivalent (e.g., local virtues will be less praiseworthy than global ones), local virtues can play many of the roles traditionally associated with global virtues. Just like global virtues, for instance, local virtues can guide moral deliberation, because they enable us to predict the behavior of ourselves and others (albeit with respect to a narrow domain). To illustrate, Doris gives the example of a spouse who has the narrow virtue (or vice) of being loyal to their spouse only when sober. Should the *loyal-only-when-sober* spouse accept their flirtatious colleague's invitation to dinner and drinks? No, says Doris: if the spouse knows they are loyal-only-when-sober, their moral deliberation should lead them to avoid situations in which they risk jeopardizing their relationship. Indeed, Doris (1998) concludes, to the extent that local character traits highlight our "situational liabilities" (p. 519), they are much better predictors of human behavior than their crude, empirically inadequate global counterparts, and therefore provide superior grounds for moral decision making.

While Doris focuses on the role of local traits in moral deliberation, other advocates of a hybrid response to situationism have used local traits to reconstruct more central features of traditional virtue theory. Peter Vranas (2005), for instance, notes that an important function of virtue and vice terminology involves evaluating someone's character *as* good or bad. But situationist research suggests that most persons are good in some contexts and bad in others; that is, their character is fragmented and therefore neither good nor bad overall but rather "indeterminate" (p. 3). Since determinate character evaluations serve important moral purposes, Vranas argues we should replace global evaluations with "*local* evaluations of people in light of their behavior in restricted ranges of situations" (p. 3). Such local evaluations would enable us to blame, say, the spouse who is prone to infidelity when under the influence of alcohol without condemning the spouse's overall character. Candace Upton (2009a) makes a similar point, arguing that local traits can facilitate accurate moral appraisals if they are based in virtuous or vicious beliefs, desires, or other mental states. This suggests that local virtues and vices can play the *evaluative* role associated with their global counterparts as well.

Combining these different strands of a hybrid response to situationism, the picture emerges of a virtue theory attuned to the complex, fragmented nature of our individual moral psychology. Instead of conceiving of virtues and vices as global traits that produce cross-situationally consistent behavior in many different circumstances, this hybrid response conceives of virtues and vices as *local* traits that manifest themselves in specific contexts. As we will now see, *group* settings are an important context of locally virtuous or vicious behavior.

3.4 The Situational Influence of Group Dynamics

Although the fundamental unit of analysis in the situationist debate so far is individual behavior, many of the situational influences revealed by psychological experiments involve group dynamics and manifest themselves primarily in social settings. Social psychologists have shown, for instance, that social pressures often lead us to conform our behavior to others within our social group (e.g., Asch, 1955); that our behavior changes depending on what role we occupy within a group (e.g., Buchanan & Huczynski, 2017; Zimbardo, 1973); and that we are less likely to help others in need if we are in the presence of other persons (Darley & Latané, 1968; Fischer et al., 2011). These findings are consistent with what many of us intuitively know: we often act differently in groups than we do when we are alone. The widespread influence of group dynamics on our behavior therefore invites us to shift our focus away from an approach to situationism that centers on isolated individuals to one that explicitly studies the impact of situational influences on individuals as they combine within groups.

If we conceive of virtues as *global* traits of character, the fact that we are liable to social pressures in group interactions may just be more evidence that we ought to abandon virtue theory altogether. To see this, consider research on the bystander effect (though one could equally turn to one of the other experimental traditions mentioned in the previous paragraph). One of the most well-established findings in all of social psychology, the bystander effect occurs when the presence of other persons has an inhibiting influence on our helping behavior (Fischer et al., 2011). In their classic study of the bystander effect, for instance, John Darley and Bibb Latané (1968) found that while 85% of study participants appeared willing to help someone in distress when alone with the victim, that rate dropped to a mere 31% as the number of bystanders increased. Similar results have been reported many

times since (see, e.g., Myers, 2008 for an overview), suggesting that most of us are less likely to engage in helpful behavior in the presence of onlookers.

This behavioral variability is seemingly at odds with the widespread existence of the global – that is, cross-situationally consistent – trait of being *helpful* or *compassionate*. After all, possessing the cross-situationally robust trait of being compassionate entails that one is disposed to helpful behavior across a broad range of circumstances. Since it is difficult to see why the number of witnesses should make a morally salient difference to our helping behavior, these circumstances no doubt include social settings such as the one simulated in bystander research. Indeed, provided it is within her means to do so, we would expect someone with a compassionate character to help someone in distress *regardless* of how many onlookers are present (unless, of course, she has reason to believe that these onlookers are better placed to provide the help required).[49] The main takeaway from bystander research, however, is that most of us consistently fail to do so. Hence, the results of bystander research strongly suggest that most of us lack the global traits – and therefore global virtues – of compassion and helpfulness.

Of course, there are various ways of accepting the conclusions of bystander research while avoiding the conclusion that we ought to abandon all virtue theory. Besides the responses surveyed in the previous section, we could turn to the philosophical literature on disposition masking (e.g., Bird, 1998; Choi, 2011; Johnston, 1992). A disposition is said to be *masked* if an external factor neutralizes its tendency to produce a characteristic type of behavior. Though glass is *disposed* to break when struck, for instance, it will not break as easily when packed in bubble wrap. The bubble wrap does not make the glass any less fragile – the glass would still break if struck without the wrap – but acts as a "masker" (Johnston, 1992, p. 233) that conditionally inhibits its tendency to break. Similarly, we could conclude that social pressures such as the ones studied in bystander research mask robust character dispositions like helpfulness and compassion. On this view, someone can genuinely be disposed to helpful behavior while subject to social influences that temporarily inhibit that behavior under certain conditions. Thus, it would be possible for someone to count as

---

[49] There are other (legitimate) reasons a compassionate person may fail to assist in the presence of onlookers. A compassionate doctor on her way to a life-saving surgery, for instance, may observe the number of witnesses and trust that one of them will help, continuing her way to the hospital. Assume, for present purposes, that no such special circumstances obtain; by and large, research on the bystander effect concerns scenarios in which the individuals involved are at least equally well-placed to provide aid.

possessing global virtues or vices even if that person did not manifest virtuous or vicious behavior under circumstances in which those dispositions were masked.

Instead of concluding that character dispositions do not exist, then, we could take the findings of bystander research as evidence that certain social pressures and group dynamics tend to mask our character dispositions. The problem with this strategy is that social psychological research has identified *so many* social influences on our behavior – so many potential *maskers* – that we should barely expect to see these character dispositions manifest themselves in characteristic types of behavior. Whatever the status of these character dispositions is, they are not *virtues*. Because unlike skills and other capacities, a trait counts as a virtue only if it is reliably exercised under relevant conditions (Foot, 1978, p. 9). In the words of Alasdair MacIntyre (1981), "[w]e cannot be genuinely courageous or truthful and be so only on occasion" (p. 185). While a skilled piano player who chooses not to play the piano very often may be no less skilled, a person who rarely performs acts of courage in situations that demand courage can hardly be called courageous (see also Meilaender, 1984). It is part of being a virtuous person that one performs virtuous actions, and it is in this regard that someone whose virtuous dispositions are masked and therefore do not issue in trait-relevant behavior falls short. What appeared to be an attractive way of accommodating bystander research thus turns out to conflict with virtue theory after all.

A more promising, *hybrid*, response to situationist research on group dynamics accepts that the influence of social pressures is at odds with the widespread existence of global character traits, but attempts to reconstruct virtue theory on the basis of *local* virtues and vices. Proponents of this strategy view the behavioral regularities observed in bystander research as the manifestation of precisely such local traits. On this view, the fact that most of us are consistently less inclined to help others when we are in the presence of bystanders is evidence that many of us possess the local vice of being *unhelpful-in-groups*. By the same token, the fact that a majority of persons appears willing to help others when they are alone with the victim suggests the widespread existence of the local virtue of being *helpful-when-alone*. Other experimental traditions, such as the ones associated with conformity research, point to yet other local virtues and vices that manifest themselves primarily in group settings. The fact that most persons conform their attitudes to those of their peers (Asch, 1955), for instance, may indicate that being *deferential-to-group-attitudes* is another widespread local vice (or virtue, depending on the context).

Once we view situationist results as the manifestation of local virtues and vices, *group settings* appear as an important context in which these local virtues and vices manifest

themselves. Indeed, many of the behavioral regularities revealed by situationist experiments involve group dynamics – a fact that is hardly surprising given the crucial role social groups play in structuring our lives. "Behavioral patterns are not, for the most part, haphazard," writes Doris (1998), "because the preponderance of our life circumstances may involve a relatively structured range of situations" (p. 508). Since group settings structure a particularly large share of the situations we encounter, it stands to reason that they are the site of a particularly large share of local virtues and vices. As we will see shortly, we can put this realization to good use when we consider questions of organizational and institutional design. Before we move on, however, it is worth pausing on Doris' contention that local traits can play many of the roles traditionally associated with global virtues and vices.

Although Doris (1998, 2002) believes that local traits are an adequate substitute for global traits in theories of virtue with respect to moral deliberation, the influence of groups on our behavior poses a *prima facie* problem for Doris. As we saw in the previous section, Doris believes that facts about our situational liabilities – such as the knowledge that one is *loyal-only-when-sober* – should figure prominently in our moral decision making. Specifically, Doris (1998) writes, "we should invest more of our energies in attending to the features of our environment that impact behavioral outcomes" (p. 515), so that we can avoid "morally dangerous situations" or what he calls "near occasions for sin" (p. 517). Just like the situationally disloyal spouse, then, the person who knows they are *unhelpful-in-groups* or *deferential-to-group-attitudes* should avoid putting themselves in situations in which they are likely to err. And if we, as bystanders, know that a majority of human beings are *unhelpful-in-groups*, this might figure in *our* moral deliberation by encouraging us to step up in emergency situations.

But while Doris' advice to avoid morally dangerous circumstances seems appropriate for the disloyal spouse, it is less useful in relation to the bystander effect and other circumstances that involve group dynamics. For though it is relatively easy to turn down a dinner invitation, it is all but impossible to avoid the circumstances in which the bystander effect is liable to occur. After all, social psychological research has established that even the presence of *one* other person tends to have an inhibiting influence on our helping behavior. Should the person who is *unhelpful-in-groups* avoid all social settings in which her helping behavior may be required? That seems absurd, particularly because it is difficult to predict when we may be called upon to help in emergency situations. Indeed, part of the reason situationism has been so worrying for virtue theorists is that we can scarcely avoid many of the situational liabilities social psychologists have identified. For most of us, heeding Doris'

call to avoid near occasions for sin might require that we live our lives, as far as possible, as hermits.

Of course, we can avoid this rather extreme conclusion if we put less emphasis on the *avoidance* of morally dangerous circumstances, and more on the importance of *recognizing* when these circumstances are likely to occur. If I *know* that I am susceptible to the bystander effect, I can attempt to mitigate its effects even if I cannot avoid social settings altogether. There is some evidence, for instance, that we can train ourselves to become "active bystanders" by preparing for bystander scenarios in advance (Staub, 2015). Indeed, just *knowing* of the bystander effect has been shown to increase the likelihood one will take action in emergencies. Insofar as becoming aware of the situational influences on our behavior enables us to diminish their unwanted effects, then, Doris is no doubt correct that recognizing these situational liabilities can make us much better moral agents.

Once we recognize that many of the morally dangerous circumstances identified by situationist research involve group settings, however, other strategies for dealing with situational influences on our behavior also suggest themselves. In particular, the powerful influence of groups on our behavior encourages us to take a closer look at *how* groups structure the range of situations individuals encounter, and at *how* these facilitate or obstruct the exercise of local virtues and vices. This entails moving away from a purely individualist perspective to one that includes the interface between individuals and groups. With this deeper understanding of the interplay between group dynamics and individual behavior in hand, we can then *design* or *engineer* group contexts to reduce the exposure to harmful situational influences and increase the exposure the beneficial ones. In other words, while Doris' preferred strategy for dealing with situationist psychology centers on what *individual moral agents* can do to mitigate their situational liabilities, this strategy involves the attempt to shape the contextual factors that engender these liabilities in order to promote virtuous conduct.

To illustrate what I have in mind, consider again research on the bystander effect. Although the presence of bystanders usually inhibits our helping behavior, social psychological experiments have identified a number of contextual factors that increase or diminish the bystander effect. For example, individuals are even less likely to intervene when they are in the presence of *powerful* bystanders, such as their supervisor or other persons in a position of authority (Ashburn-Nardo, Blanchar, Petersson, Morris, & Goodwin, 2014). By contrast, individuals are more likely to intervene when they are in a position of authority themselves, when they perceive their environment as supportive of intervention, or when

others in their environment have previously exhibited helping behavior in similar situations (Burn, 2009; Hussain, Shu, Tangirala, & Ekkirala, 2018; Mudde, Hoefnagels, Wijnen, & Kremers, 2007; Sanderson, 2020). Since individuals typically have little control over whether they find themselves in such supportive environments and cannot always avoid powerful bystanders, there is only so much they can do to follow Doris' advice to avoid morally dangerous circumstances. Group leaders and other persons involved in creating and managing social groups, however, often do have considerable powers to shape such contextual factors.

Indeed, empirical research suggests numerous ways in which those in charge of managing or designing organizations and other social groups can "nudge" group members into virtuous behavioral patterns (Sunstein, 2014). Hussain et al. (2018) demonstrate this in relation to the bystander effect when they discuss the problem of "open secrets" in organizations. Why, they ask, do an organization's members so often fail to speak up despite it being an open secret that certain problems persist within the organization? Consider, for instance, the case of Harvey Weinstein, whose affiliates reportedly knew of his sexual transgressions years before they erupted into public scandal. Hussain et al. diagnose this as a classic instance of the bystander effect: the fact that many persons not only know of a certain problem, but also know that *others* know, diffuses the sense of responsibility these persons have for addressing that problem, inhibiting their helping behavior. But, Hussain et al. argue, managers can reverse this dynamic by fostering an environment that is supportive of intervention. In practice, this means leading by example (making sure that group leaders and others in positions of authority speak up about perceived problems themselves), cultivating an organizational culture that encourages group members to voice their concerns, and rewarding whistleblowers. Each of these contextual variables can bend a general tendency *not* to intervene when bystanders are present into a general tendency to intervene when required.

Moreover, if the hybrid response is correct in viewing situationist behavioral regularities as the manifestation of genuine virtues and vices, albeit local ones, group leadership can thus play a crucial role in promoting locally virtuous (and diminishing locally vicious) behavior by directing the situational influences to which group members are exposed. This point generalizes to situational influences that are not purely social in nature. As I show in the next chapter, for instance, organizations should avoid pressuring their members to perform critical tasks in a hurry.[50] Not only does being in a hurry make one less likely to engage in helping behavior, as the Princeton Theological Seminary experiment demonstrates,

---

[50] See de Bruin (2015) for a similar argument in relation to the financial crisis.

being in hurry can also impede such epistemic virtues as diligence and care. The next chapter illustrates this by way of a case study of Boeing, whose ill-fated 737 Max planes were involved in two deadly and widely publicized crashes in 2018 and 2019. Though the exact cause of the crashes remains under investigation, it is clear that Boeing designed the planes under severe competitive pressures. As we will see, these pressures culminated in an organizational culture in which situational influences produced "rushed engineers" delivering "sloppy blueprints" (Gelles, Kitroeff, Nicas, & Ruiz, 2019, para. 29). Had Boeing leadership fostered a less rushed, more supportive environment, its engineers would have had a better opportunity to exercise their (epistemic) virtues in designing the 737 Max.

To sum up, a hybrid approach to situationism opens up a view not yet present in the literature. Rather than viewing situationism as an embarrassment for virtue theory, we could use its insights to harness locally virtuous behavior by anticipating and working around the situational liabilities to which we are exposed. Given that organizations and other social groups structure many of the situations we encounter, it is especially important that we keep situationist research in mind when we engage in organizational design, ensuring, to the extent possible, that groups expose their members to situational influences conducive to virtuous behavioral patterns and unfavorable to vicious ones.

3.5 Expanding the Situationist Challenge to Group Agents

In the previous sections, we have seen that the situationist challenge hardly presents a knockdown argument to virtue theory in all of its many forms. While situationists have forced virtue theorists to contend with recent psychological research, virtue theorists have responded with convincing conceptual, empirical, and hybrid defenses of the virtue theoretical project. As a result, contemporary contributions to virtue ethics and epistemology are more attuned to the psychological complexity of individual human agents than ever before. Although this focus on individual human agents is understandable, we routinely apply virtue theoretical insights to group agents as well. In the remainder of this chapter, I therefore expand the situationist challenge in new directions by critically examining the implications of situationist critiques for group agents. Analyzing the situationist challenge at the collective level, the focal point of my discussion will be the idea, defended at length in the previous chapter, that we can often characterize the collective character of group agents as virtuous or vicious in its own right.

Indeed, a growing body of literature argues that group agents such as corporations and other organizations can realize virtuous or vicious character traits at the collective level (de Bruin, 2015; Fricker, 2010, 2013, 2020; Konzelmann Ziv, 2012; Lahroodi, 2007; Sandin, 2007). According to this literature, group agents can instantiate such collective virtues and vices as open-mindedness, courage, and greed independently of the character of individual group members. Here, then, is a natural thought. Whereas our individual psychologies are subject to certain biological constraints, the cognitive infrastructure of group agents is realized through various decision-making mechanisms, organizational structures, and other information processing procedures that operate free from the constraints of human biology (List & Pettit, 2011; Tollefsen, 2015; Tuomela, 2013).[51] Since group agents do not face the same psychological limitations as individual agents, they are likely immune from the situational influences that impinge on individual cognition. Hence, group agents do not have to worry about the kind of situationist concerns levelled against individuals, rendering the point of expanding the situationist challenge to group agents moot.

Though this is an attractive line of thought, I argue below that is ultimately a mistake to hold that situationist critiques gain no traction on group agents whatsoever. While the sorts of situational influences that destabilize virtuous or vicious dispositions at the collective level may be different from those that affect individuals, there is ample evidence that groups, too, are vulnerable to outside pressures. These pressures will not, as in the case of individuals, consist in ambient sounds, smells, or other contextual factors that involve distinctly human capabilities, but rather consist in situational liabilities that impact those capabilities realizable by group agents. This asymmetry between the sorts of situational influences that affect individual agents and those that affect group agents opens up a crucial challenge. In order to expand situationist critiques to group agents, we must find out what situational dynamics have the potential to dislodge *collective* virtues and vices. Since this is relatively unexplored terrain, a new situationist program is called for that studies the behavioral variability of collectives under a range of different situational influences. I sketch the contours of that program in what follows, paving the way for future research.

To get started, we should consider what a situationist challenge to group character might look like. Just as individual character traits, collective character traits are cross-situationally robust dispositions to characteristic types of behavior (Fricker, 2020; Lahroodi,

---

[51] Except, of course, to the degree that these mechanisms, structures, and procedures are implemented by individual agents. Since this section primarily addresses traits realized at the collective level, we can bracket that question here.

2007). These dispositions are such that they issue in trait-relevant behavior across a broad range of relevant circumstances. An organization with the character virtue of justice, for instance, is *disposed* towards just conduct in situations that call for fairness, respect, impartiality, and so on. It would count *against* the possession of such character traits if these dispositions were easily destabilized by situational influences irrelevant to the trait in question. Thus, say, a police force would not count as a just organization if it only exhibited just behavior under public scrutiny, or if it threw justice in the wind to meet ambitious ticket or arrest quotas resulting from political pressures (Ossei-Owusu, 2021).

For the situationist challenge to get a hold on group agents, however, it is not sufficient for *some* groups to be vulnerable to situational influences in this way. After all, the fact that some groups fail to realize cross-situationally robust character traits due to situational influences does not problematize the notion of group character as such. Even if it is true that some police forces buckle under political pressures, this does not mean it is impossible for other police forces to resist these pressures. Situationists would mount a more convincing challenge to the notion of group character if, as they have attempted to do in the case of individuals, they could show that a significant number of groups are vulnerable to trivial situational influences. This challenge would be more convincing still if these influences turned out to be as numerous as the factors that influence individual behavior. Only if situationists can show that group agents are vulnerable to situational influences across the board do they manage to call into question the empirical basis for believing there is such a thing as *group character* at all. The burden, then, is on situationists to show not just that group agents are vulnerable to situational influences, but that these vulnerabilities and the influences that trigger them are relatively widespread.

Though research in this area is not as developed as the social psychological research that problematizes the notion of individual character, there is empirical evidence that may suggest groups often lack the cross-situational consistency required for the possession of virtuous or vicious character traits. This evidence has been most pronounced in the study of the so-called *home advantage* in professional sports (Boudreaux, Sanders, & Walia, 2017; Courneya & Carron, 1991; Nevill & Holder, 1999; Pollard, 2008). A widely studied phenomenon in sports and behavioral economics, the home advantage consists in the edge home teams have over visiting teams in sport competitions. Observed in a large variety of team sports, the overall home team advantage can be significant, elevating the likelihood of winning by 20.4% in the NBA (Boudreaux et al., 2017) and a global average of 61.5%(!) in professional soccer (Pollard, 2006). While the dynamics that give rise to the home advantage

are complex, there is broad consensus among experts that such situational influences as the presence of a supportive crowd play a crucial role. In their study of the home advantage in professional basketball, Boudreaux et al. (2017) even suggest that biased crowd effects may explain the home advantage almost entirely.

We should not expect to see such behavioral variability if team performance were wholly a function of cross-situationally robust skills or traits realized at the collective level. Cross-situational consistency demands that trait-relevant behavior manifests itself across a broad range of trait-relevant conditions. Clearly, these conditions include both home and away games, in front of supportive and hostile audiences, in the case of professional sports. Just as it should not make much of a difference to a compassionate person whether she is in a hurry or not, it should not matter to, say, a skilled soccer team whether it plays in its home stadium or at an away field. Nor, on the face of it, would the absence or presence of a supportive crowd seem to bear much on the virtues required of a successful soccer team. On the contrary, one would expect an excellent team to rise above such contextual factors. Plainly, however, these situational factors *do* make difference. This seems to suggest, then, that at least *some* collectives are vulnerable to situational influences in a way that counts against the possession of robust character traits.

As stated before, the fact that some collectives are easily influenced by situational factors far from proves that the notion of group character lacks empirical support. But it *does* problematize the notion that situationist critiques have no purchase on group agents whatsoever. At this point, however, one may object to my analysis of the literature on home advantage effects in two ways. First, one may be skeptical that the situational liabilities identified by research on the home advantage are irreducibly collective at all. Rather, one may suspect the home advantage effect is caused by situational influences on *individual players*. If that is correct, research on the home advantage effect belongs to the individual level of the situationist program, and not the collective level I aim to develop here. Second, one may worry that professional sports teams do not possess group *agency* in the first place. While sports teams consist of players who coordinate their actions in order to achieve a common goal, they arguably lack the highly complex, integrated information processing and decision-making capabilities that allow such canonical group agents as corporations, organizations, and state institutions to speak and act as *one*.

The first of these objections seems unlikely in light of empirical research on the home advantage effect in individual sports. While a home advantage has been observed in subjectively evaluated individual sports ranging from diving, gymnastics, and figure skating

to boxing, the home advantage is notably absent in such objectively evaluated sports as individual tennis and golf (Balmer, Nevill, & Lane, 2005; Jones, 2013; Nevill, Holder, Bardsley, Calvert, & Jones, 1997). The fact that the home advantage only seems significant in individual sports that are evaluated at the discretion of a judge has led some experts to suggest that these results are not due to situational factors such as the presence of a supportive audience, but rather due to referee bias (Balmer, Nevill, & Williams, 2001; Balmer, Nevill, & Williams, 2003). Though this does not preclude the possibility that the home advantage effect observed in team sports is caused by individual situational liabilities, that does seem improbable in view of the absence of such liabilities in individual sports.[52] Hence, it seems appropriate to analyze the home advantage observed in team sports in terms of group dynamics manifested at the collective level after all.

The objection that professional sports teams do not meet the conditions for group agency, however, is slightly more difficult to diffuse. The sting of this objection depends partly on one's preferred account of what these conditions are. According to Margaret Gilbert (1989, 2013), for instance, groups possess agency to the extent that its members are jointly committed to acting *as one*. Insofar as the players on a soccer team have jointly committed to play together as one, then, it arguably comes out as a group agent on Gilbert's account. But it may appear that a soccer team does not meet the conditions for group agency imposed by more stringent accounts, such as the one defended by Christian List and Philip Pettit (2011). For List and Pettit, group agency requires complex, integrated information processing and decision-making capabilities that enable the group to act as an autonomous agent. This entails the capacity to formulate and evaluate beliefs at the collective level, which one may suspect sport teams lack. Indeed, one might think that while individual players have certain beliefs about, say, the state of play in a soccer game, the players typically lack decision-making mechanisms such that we can ascribe beliefs to *the team* as such (that is, *qua* group).

Although the devil is in the details, I suspect that soccer teams ultimately *do* come out as group agents on List and Pettit's account as well. After all, teams often do decide what strategy to pursue *as a team*, with mechanisms in place to ensure that collective intentions are formulated and carried out. The intention to, say, play more aggressively in the second half is communicated by the coach during halftime, and enforced by the captain, with all players

---

[52] One may suspect that the home advantage effect is absent because individual sporters do not have the same practical possibilities of benefiting from a home advantage (e.g., insofar as they lack a home stadium). However, research on the home advantage effect has tried to control for this by focusing on Olympic events where sporters performed in their home country and in front of supportive crowds.

aware that the coach and captain together decide what the *team* does. Still, I take it that there are meaningful differences between soccer teams (considered *qua* squad) and organizations that implement group agency on a larger scale. The situational influences that impact a soccer team playing on pitch are likely different than those that affect a transnational corporation with offices around the world. We have reason, then, to consider the implications of situationist critiques for corporations, organizations, and other collectives that implement group agency on a larger scale.

In one of the few academic discussions of this topic so far, Per Sandin (2007) argues that incorporated agents such as businesses and other organizations likely avoid the situationist challenge because they are "significantly less fickle than individuals" (p. 306). To substantiate this, Sandin draws on Peter French's (1995) work on corporate responsibility. Corporations, French writes, "will probably prove more capable of consistent and dependable ethical behavior than humans" (p. 90) in virtue of their "constitutive structures" (p. 85). These structures, which are typically fixed by corporate charters and other official documents, determine the mechanisms by which corporate decisions are made and establish the roles, rights, and duties of members of the organization (French, 1979). Corporate structures are usually designed to provide a high degree of stability, ensuring that the corporation survives changes in membership, remains focused on its objectives, and is capable of responding to changing (market) conditions. Consequently, corporations possess a high degree of autonomy over and above individual members and so meet List and Pettit's criteria for group agency as well.

While Sandin's remarks are instructive, however, they do not show that corporations are immune from situational influences in the way that is required for the possession of robust character traits. As we have already seen, the possession of robust character traits requires a *particular type* of stability, one that manifests itself across a broad range of situations. In other words, it is not sufficient for organizations to exhibit behavioral consistency under particular circumstances. But for all Sandin and French have said about constitutive structures, it is at least conceivable that they only produce consistent behavior with respect to a narrow domain. Indeed, corporations are typically set up such that they are effective in realizing the objectives set by group leadership, often anchored in corporate charters as well, but that efficiency need not translate to other areas. Thus, while corporate structures typically protect the corporation from the whims (and potential misconduct) of individual members, with protocols and safeguards ensuring the corporation stays on-task, they may not protect the organization in unforeseen circumstances. Sandin may be correct that corporations are more

stable than individuals in *some* respects (i.e., they are not vulnerable to the whims of one person), but they may nevertheless be vulnerable to situational influences in others. The task, then, will be to find out if, and in what respects, corporations *are* liable to situational influences.

I have no doubt that more empirical research is required to meet this challenge, but here is some *prima facie* evidence that suggests Sandin's conclusion that corporations may avoid the situationist challenge is not as straightforward as it seems. Consider, again, the case of Boeing, which I discuss in detail in the chapter that follows. Earlier, I suggested that Boeing's engineers were unable to exercise their epistemic virtues due to various competitive pressures. The problem, however, affected not just Boeing's engineers, but the organization *as a whole* (The Boeing 737 MAX Aircraft: Preliminary Investigative Findings, 2020; Gelles et al., 2019).

One of the largest plane manufacturers in the world, Boeing aims to make a profit designing, selling, and maintaining high-end aircraft. Since these are all time- and resource-intensive tasks, Boeing's constitutive structure is ideally suited to the pursuit of long-term projects. For the most part, that structure has served the company well, providing the consistency and stability required to realize its ambitions across the span of multiple years or even decades. When its main rival Airbus surprised Boeing with the success of its new airplane in 2011, however, Boeing was under tremendous pressure to update its airplane designs *in the short term*. To *this* end Boeing's constitutive structure was not conducive. In the hurry to catch up, Boeing *the corporation* made a series of poor decisions at the collective level, creating the conditions leading up to the crashes of two of its airplanes (Laker et al., 2021). In spite of its robust constitutive structure, this seems to suggest Boeing was vulnerable to the collective equivalent of the dynamic under investigation in the Princeton Theological Seminary study. Just like individuals, Boeing's character dispositions (in this case, to exercise diligence in the design of new aircraft) were easily dislodged when it found itself in a rush.

If the troubles at Boeing were unique in the world of business, they would hardly present a problem for the notion that groups can realize cross-situationally robust character traits. But in fact, many organizations have suffered similar fates. Enron, the American energy trader, was a paragon of corporate virtue until the pressure to meet increasingly unrealistic targets caused it to collapse in a spectacular bout of corruption (Sims & Brinkmann, 2003). More recently, the German car company Volkswagen – also once lauded for its corporate ethics – became embroiled in a widely publicized emissions scandal, which is due at least in

part to competitive pressures in the U.S. market (Bovens, 2016; Rhodes, 2016). Though these were each companies whose constitutive structures provided the stability required to grow into large multinational organizations, these structures did not protect them from the disruptive influence of market pressures. One may even suspect rigid constitutive structures were part of the problem, rendering these organizations inflexible and making them slow to adapt to changing circumstances.

Corporations, it would seem, are by no means immune from situational influences. This suggests that the situationist challenge extends beyond individuals to group agents as well. However, that does not mean that we should abandon the concepts of group virtue and vice. Just as virtue theorists have done in the case of *individual* character, we can push back against the claim that situationism undermines the notion of *group* character. In line with the hybrid response developed in section three, one possible strategy for doing so immediately suggests itself. That is, we could reject the notion that collective virtues and vices require cross-situational consistency across a broad range of relevant circumstances. Instead, we could argue that *local* behavioral regularities are a sufficient basis for attributing virtues and vices to a group. Collective virtues and vices of a local kind would then speak to the group's qualities under specific circumstances.

Indeed, a hybrid response to situationism at the level of collectives is particularly attractive in view of the narrow ends groups typically serve. One reason for insisting on the importance of cross-situational consistency in the case of individual virtue and vice is that human flourishing plausibly involves excellences across many different domains. On Aristotle's account of virtue, the flourishing agent excels as a friend, warrior, and citizen – all roles that they may assume at one point or another. But an important disanalogy between individual and group agents is that the agency of groups is often limited to specific domains (Konzelmann Ziv, 2012). For instance, many jurisdictions restrict the agency of incorporated groups to the objectives explicitly stated in their articles of association. Since the objectives of a company like Boeing are to further aerospace innovation and design and sell aircraft, we should not expect it to excel in other domains. The same goes for *un*incorporated groups that serve specific ends. A parent-teacher committee, for example, serves important purposes in local school districts, but we should not expect it to display competence in other areas.

Since group agents often operate in specific domains then, local virtues and vices that are indexed to these domains are more than capable of capturing the relevant senses in which these groups can excel or fall short. But even if one insists that collective virtues and vices require cross-situational consistency, the concerns sketched above should not lead us to reject

the notion of group character. Supposing it *is* impossible for groups to manifest cross-situational consistency, this does not mean that collective virtues cannot serve as regulatory ideals. In the domain of individual virtue, for instance, Mark Alfano (2013) has argued that we should aspire to be virtuous *even if* virtue represents an unattainable ideal. That may hold for collective virtues as well. While groups may never be able to stave off all situational liabilities, *aspiring* to cross-situational consistency may incentivize groups to perform as well as they possibly can.

As it stands, however, it is too early to conclude that these situational liabilities are sufficiently widespread that the notion of group character has no purchase whatsoever. For all I have said, the situational liabilities of sports teams and organizations like Boeing, Enron, and Volkswagen can be overcome. To answer that question definitively, more empirical work is needed on the behavioral variability of group agents under various kinds of external pressures. Additional research in this area will likely pay off in more ways than one. Once we have a better understanding of the dynamics that influence character dispositions at the collective level, we are also in a better position to arm ourselves against them. This is reflected in the growing body of literature on organizational resilience, which studies what organizations can do to improve stability in times of crises and uncertainty from a management perspective (Suarez & Montes, 2019; Vogus & Sutcliffe, 2007). Ultimately, I am optimistic this research will reveal that at least some group agents are resilient enough to warrant the label of (collectively) *virtuous* or *vicious* even in a cross-situationally robust sense. Just as in the case of individuals, the situationist challenge to *group* character does not render the notion of group character useless.

3.6 Conclusion

Although virtue theory has grown into a vibrant field, it has also attracted its fair share of critics. The most damaging critiques have focused on virtue theory's empirical commitments, targeting the notion that individuals are capable of realizing the cross-situationally robust character traits that virtue theory presupposes. As we have seen, these critiques have not succeeded in establishing that we ought to abandon virtue theory altogether, though they have forced virtue theorists to take seriously the empirical commitments of their theories. While the vast majority of situationist critics adopt the perspective of individual human agents, this chapter has sought to extend the situationist challenge in two ways. The result is a situationist

program characterized by three levels of analysis. The first level of analysis examines the impact of situational influences on individuals in isolation; the second studies the impact of situational influences at the interface of individuals and groups; and the third level of analysis studies the impact of situational influences on groups and collectives *as such*.

While situationist research at the individual level has been underway for some time, research at the other two levels is relatively scarce. This means that more research is required to map the situational influences that can destabilize virtuous or vicious dispositions within groups and at the collective level. As I have shown in this chapter, we can use research on the situational influences that affect the interface between individuals and group dynamics to promote the exercise of *local* virtues in group members while diminishing local vices. Research on the situational influences that affect the behavior of groups *as such* (that is, *qua* group) may enhance our understanding of the sorts of mechanisms required to promote group- and organizational resilience at the collective level. Though this chapter marks an important first step towards a multifaceted approach to situationism that takes seriously the collective aspects of our behavior, it leaves a variety of questions unanswered. I answer some of these questions in the chapter that follows, which is a case study of the dynamics, situational and others, that have disrupted the character dispositions of Boeing and its employees.

Chapter 4 – Collective Epistemic Virtue and Vice: The Case of Boeing[*]

4.1 Introduction

Moments before Lion Air Flight 610 crashed into the Java Sea with dramatic loss of life, its pilots frantically searched the jet's handbook in an effort to understand why its nose suddenly pushed down. Working through checklist after checklist amid a growing number of alarms, they never found the information. An official investigation into the causes of the crash later determined that the pilots could never have found what they were looking for. Boeing, which produces the 737 Max jet the pilots were flying, had deliberately omitted crucial information about the flight control system from the manual—a decision that turned out to be part of wide-ranging cost-cutting measures (Komite Nasional Keselamatan Transportasi, 2019). Five months after the Lion Air crash, the pilots of an Ethiopian Airlines 737 Max jet saw the nose of their plane unexpectedly push down as well, and they, too, were unable to find relevant information about the flight control system. Like the Lion Air flight, their plane crashed moments later.

Even before these two disasters, 737 Max pilots had complained that they "lacked the knowledge" (para. 23) to operate the plane (Fallows, 2019). It has since emerged that insufficient documentation is only the tip of the iceberg. The tale of the 737 Max crashes is a story of epistemic vice through and through. At the center of that story is Boeing, the world's largest aerospace company, whose "culture of concealment" (p. 3) is currently the target of an investigation led by the U.S. House Committee on Transportation and Infrastructure ("House Committee" hereafter) (House Committee on Transportation and Infrastructure, 2020). In a report of its preliminary findings, the House Committee accuses Boeing not just of failing to

---

provide pilots with proper documentation, but also of withholding "crucial information" (p. 3) from customers and federal regulators. Internal communications obtained by the committee further reveal that Boeing ignored several whistleblower complaints and safety warnings issued by its technical pilots and quality assurance officers. Aviation analysts have in fact described Boeing's attitude leading up to the 737 Max crashes as "arrogant" (Gelles, Kitroeff, Nicas, and Ruiz, 2019, para. 21), and while they do not use the term, from what they write it is clear that this arrogance is at least partly epistemic (Laker, Cobb, and Trehan, 2021).

Epistemic arrogance is a canonical epistemic vice: a character trait that obstructs the realization of such epistemic goods as knowledge, understanding, and wisdom.[53] While the concept of epistemic vice has proved useful in analyzing the epistemic state of individuals, philosophers have only recently begun to apply it to *collective* entities such as work teams, boards of directors, committees, or even to entire corporations. We think this move is fully justified by the role collectives play in the economy of knowledge and belief. They are not just the site of beliefs in their own right, as when we say that Boeing believes its planes to be safe. More than that, they are key in the *transmission* of beliefs, as when a pilot relies on Boeing for proper documentation.

But what does it mean to say that Boeing was *arrogant*?

This chapter critically examines extant theorizations of collective epistemic virtue and vice. It lays out certain conceptual problems and proposes ways of overcoming these problems. Since this chapter draws liberally on the functionalist account of collective virtue and vice developed in chapter two, some overlap is inevitable. However, this chapter aims to extend the findings of chapter two explicitly to *epistemic* virtues and vices. Thus, it argues for what could be called a *functionalist* account of collective epistemic virtue according to which epistemically virtuous groups are organized so as to function *as* epistemically virtuous agents. On the view defended here, an epistemically virtuous organization typically has three components: it exhibits *organizational support for virtue*; it has *organizational remedies against vice* in place; and *it matches the individual virtues of employees to the organization's functions*, for instance in hiring decisions. Organizations may manifest collective epistemic vice if they fail to enact a corporate structure that is virtuous in this way. One key objective of this chapter is to illustrate the practical *real-life relevance* of an approach to collective virtue

---

[53] This chapter does not mean to take a firm position on the nature of epistemic vice. That is why, for present purposes, it adopts this rather broad characterization of epistemic vice that is loosely based on Cassam's (2019) definition of epistemic vice.

epistemology, which is also conceptually and empirically sound. The chapter therefore spends considerable time investigating Boeing's *epistemic corporate culture* (De Bruin, 2020).

4.2 Collective Epistemic Virtue and Vice

Epistemic virtues are features that make us excellent *qua* producers and consumers of epistemic goods, such as knowledge, understanding, or wisdom. By contrast, epistemic vices obstruct the realization of these epistemic goods. Virtue epistemologists disagree somewhat over the nature of these features. For virtue reliabilists such as Ernest Sosa (2007) and John Greco (2010), epistemic virtues comprise *all* stable dispositions that reliably produce true beliefs. Prime examples of reliabilist virtues include such cognitive faculties as sense perception and reliable memory. Virtue responsibilists such as Lorraine Code (1987) and Linda Zagzebski (1996), on the other hand, characterize epistemic virtues primarily as the *character traits* that mark an excellent knower. On this picture, an epistemically virtuous knower not only reliably forms true beliefs, but also cultivates such epistemically virtuous character traits as honesty, open-mindedness, and intellectual courage. Cultivating these traits requires that we are moved by virtuous epistemic *motives*, such as love of wisdom.

While the literature on epistemic vices has only recently emerged (see, e.g., Baehr, 2010; Battaly, 2014, 2016; Cassam, 2016, 2019; De Bruin, 2015), they are typically conceived of as the inverse of epistemic virtues. Thus, epistemic vices may include such unreliable faculties as poor vision and obstructive character traits as closed-mindedness, overconfidence, or hubris.

We not only attribute epistemic virtues and vices to individuals, we also regularly attribute them to groups. For instance, we commend the International Consortium of Investigative Journalists (ICIJ) for its display of intellectual courage, or we reproach the financial service providers whose misconduct the consortium unmasked as careless and dishonest. As intuitive as we find it to use the language of virtue and vice to talk about groups, we have seen in chapter two that the metaphysical status of these attributions is far from straightforward. Are we merely using a linguistic shortcut to talk about the features of its members, or do we say that the group exhibits these features *qua* group?

In chapter two, we saw that *summativists* are poised to answer that group features reduce to individual features, and so that groups lack these features as subjects in their own

right.[54] On the basic summativist analysis we considered, group G exhibits virtue or vice V only if a sufficient number of its members exhibit V. If the ICIJ consortium is a courageous institution, summativists maintain, this is because its *individual journalists* display the virtue of courage (in speaking truth to power, say). Since this entails that only individuals can be the proper subject of virtues and vices, summativists hold that the most we can do when we attribute these traits to groups is make summary reference to the traits of its individual members. We would be mistaken if we believed, for instance, that the ICIJ is the seat of courage as a subject in and of itself.

Summativism enjoys a great deal of initial plausibility. It meshes well with the widespread conviction that individual agents are the basic explanatory units of all social phenomena (the doctrine of *methodological individualism*). Moreover, it is clearly correct as an account of at least *some* collective virtues; for when we praise our group of students for their diligent work ethic, we really do appear to praise the character of our individual students.

Yet summativism often oversimplifies the relation between a group and its members. Consider Reza Lahroodi's (2007) example of a group that is collectively narrow-minded even though it is for the most part composed of individually open-minded members, which I discussed in chapter two as well. For present purposes, we could think of the board of directors of an aerospace company. As individuals, these board members are open-minded about such things as aerospace innovation, and they are disposed to give reasonable innovative ideas a fair hearing. Collectively, however, the board is not so disposed. The board believes the company's market position is sufficiently secure not to invest in innovation, and so when innovation comes up as a topic in the boardroom, it often dismisses these ideas without giving them fair consideration.

Lahroodi's thought experiment is construed in such a way that if we were to tally the number of individually narrow-minded directors, we would come up empty. He observes that a summativist should conclude that the board is not collectively narrow-minded either (that is, that the board is not narrow-minded *qua* group). This, he claims, seems wrong, as the board routinely rejects innovative ideas out of hand. Lahroodi therefore contends that summativism

---

[54] The term "summative" (p. 9) is due to Quinton (1976). See Gilbert (1989) for discussion. While Fricker (2010), Lackey (2016), and Gilbert (1989) prefer the term *summativism*, Lahroodi (2007) uses *individualism*, and List and Pettit (2011), *eliminativism*, all with subtle distinctions. We use *summativism* without privileging any of the extant views.

is incorrect as a general account of group epistemic virtue and vice: we cannot always analyze such traits as mere sums of the traits possessed by individuals.[55]

Perhaps a sophisticated summativist can account for Lahroodi's example, for it may not be individual *virtues* we should tally, but other individual features. Yet a growing number of philosophers take examples of Lahroodi's kind to motivate the search for *non-summativist* accounts of collective virtue and vice (Fricker, 2010; Jones, 2007). Non-summativists claim that the members of a group sometimes interact in such a way that they form collective agents whose properties are distinct from the properties exhibited by these members themselves.[56] In other words, they hold that groups can be more than, or at least different from, the sum of their parts, as Lahroodi's example illustrates.

A leading non-summativist account of group agency is due to Margaret Gilbert (1989, 2013). According to Gilbert, some groups form what she calls *plural subjects*, with intentions, beliefs, and other agential features of their own. These plural subjects are instantiated when two or more individuals jointly and openly commit to upholding these features *as a body* (Gilbert, 2013, p. 32). The journalist members of the ICIJ, for instance, form a plural subject of the intention to uncover fraud to the extent that they jointly and openly commit to investigate fraud and money laundering as a body, or as one. Gilbert seems to intend her notion of doing something as a body, or as one, to be read metaphorically, as she does not believe that plural subjects *literally* possess a body of their own. What she thinks is that the parties to a plural subject coordinate their actions so as to emulate a single body; hence the spirit of methodological individualism is preserved.

Gilbert (2013) is clear that plural subjects are irreducibly collective entities, because the constitutive joint commitments do not reduce to *personal* commitments. This opens up a logical space in which these commitments diverge: you can be jointly committed to narrow-minded practices, even if you are personally committed to being open-minded, just as in our example. The technical details of Gilbert's argument need not detain us here, but it may be helpful to point out that the difference between these two types of commitment is brought out by normative expectations accompanying them. When you personally commit to something, you can unilaterally rescind the commitment whenever you like. You do not owe it to anyone to follow through on your commitment. But if you committed to something jointly with

---

[55] Lahroodi's case amounts to a virtue epistemological version of what Lackey (2016) calls "divergence arguments" (p. 343).
[56] Lahroodi (2007) uses the term *anti-correlativism* instead of *non-summativism*. List and Pettit (2011) prefer *realism about group agency*, with subtle distinctions.

others, you and the other members of a resulting plural subject incur obligations towards each other. No member has the standing to rescind these joint commitments unilaterally; they can only be rescinded (without violating social norms) if everyone agrees. Joint commitments are, that is, intrinsically other-involving, and this is why we cannot perform a summative reduction of these commitments to personal ones.

While Gilbert's account can help us understand how groups could instantiate properties their members lack individually, Lahroodi doubts that it provides a fully viable model for collective virtue and vice. The problem, he claims, lies in Gilbert's requirement that joint commitments be *open*, or *transparent*, to all parties involved. Gilbert indeed holds that one of the prerequisites for being jointly committed to something is that the members of the plural subject have expressed to each other their willingness to be so committed, thereby signaling acceptance of the content of the commitment in question. As Lahroodi argues, however, this lacks plausibility when it comes to virtue and vice. A group can be open-minded, he thinks, even if its members do not know that it is open-minded, let alone have openly committed to open-mindedness; what matters is only its disposition to give a fair hearing to contrary ideas.

There are various ways of responding to Lahroodi's concerns. In chapter two, I discussed Miranda Fricker's (2010) response. Fricker argues that there is no special philosophical puzzle in holding that plural subjects can manifest epistemic virtues or vices none of its members are aware of. They may simply fail to know that the traits to which they have jointly committed count as virtuous or vicious. Just as some individuals manifest the virtue of modesty without knowing that they could be adequately described as *modest*, so the members of a plural subject may be jointly committed to routines, values, or procedures without knowing that these features constitute virtuous or vicious traits. The directors of our aviation company, for instance, need not be aware of having committed to *narrow-mindedness*, under that description. Our version of Lahroodi's example is more plausibly construed as involving a positive commitment to maintaining the corporation's legacy, which happens to have the unhappy consequence of reducing investments in research and development.[57]

---

[57] This construal of Lahroodi's case is meant to be plausible. Anticipating the next section: Boeing has reportedly resisted fully embracing computerized flight control technology for decades, believing pilots prefer to be in charge at all times. Its main competitor Airbus has installed extensive computer technology in aircraft since the early 1980s (Gelles et al., 2019).

Still, various problems with a plural subject approach remain.[58] For one, it is unclear whether Fricker's response will satisfy critics such as Lahroodi. While Fricker is arguably correct that the members of a plural subject need not construe the trait they have jointly committed to as virtuous or vicious for it to have the relevant virtue or vice, they do, on a Gilbertian analysis, need to have somehow expressed a willingness to be committed to the trait in question. You might think that even this requirement is too strong. In Lahroodi's words, it simply does not seem totally felicitous to claim that members "have to jointly accept to exercise a trait for the group to have a trait" (p. 292). There are many groups that, on the face of it, exhibit traits that their members have not accepted to exercise. Negligence may be a case in point, if it is thought that a reluctance to accept *any* commitments towards safety and diligence makes a group negligent.

Another issue concerns the empirical adequacy of a plural subject approach. One common objection to Gilbert's plural subject account is that it does not seem suitable for analyzing the agency of large groups whose members do not always know each other (e.g., Ritchie, 2020).[59] We often attribute virtues and vices to universities, multinational corporations, NGOs, and other large collectives that may be composed of thousands of members, many of whom will never interact. Boeing, for instance, has over 150,000 employees across 65 countries. It has three business divisions, and dozens of offices and manufacturing plants. Supposing that Boeing suffers the vice of arrogance, as some analysts maintain, is it really plausible to claim that its employees have all jointly expressed to each other a readiness to commit to arrogant qualities?

Given these concerns, we may want to turn to an alternative non-summativist analysis of group agency. In line with the argument developed in chapter two, a *functionalist* analysis of group agency provides a way of modeling collective epistemic virtues and vices that is particularly promising if you are driven by real-life practical concerns. On a functionalist analysis, groups possess agency insofar as they are systems that *function* as agents.[60] Christian List and Philip Pettit (2011) illustrate this using a classic belief–desire model of agency. An agent, List and Pettit argue, is a system that exhibits three features: it has beliefs about what the world is like; it has desires as to how the world should be; and it has the capacity to act on

---

[58] For a critical discussion of Fricker, see Byerly and Byerly (2016), Cordell (2017), and Konzelmann Ziv (2012).

[59] Though Gilbert (2006) aims to defend plural subject theory against chis charge.

[60] Many functionalists believe that functioning as an agent simply is what it means to be an agent. Similarly, we believe that groups that function in a virtuous or vicious way really do have group-level virtues or vices – they are not simply *as-if* virtues or vices.

these desires on the basis of these beliefs. Individual human persons satisfy these conditions, but so do many other systems, including robots, animals, and, List and Pettit maintain, some groups. A business organization, for instance, typically has desires (say, to maximize profits), beliefs (say, about market conditions), and the capacity to produce and sell goods or services on the market in order to realize these desires.

According to List and Pettit, the relationship between a group and its members is one of supervenience, so there cannot be a difference *qua* group-level beliefs, desires, and other features without there being a difference *qua* features possessed by individual group members (taking the procedure by means of which these individual features are *aggregated* into account). Accordingly, functionalism does not militate against methodological individualism. Crucially, List and Pettit argue that the members of a group can realize group agency in various configurations. There are many conceivable aggregation functions taking us from individual to group beliefs and desires, and numerous ways in which groups could act on these beliefs and desires. A group could use majoritarian voting methods, for instance, but it could also choose to adopt a dictatorial Chief Executive Officer (CEO).[61]

The crux of List and Pettit's non-summativism lies in the multiple realizability of group-level features. On a functionalist analysis, the same group beliefs and desires can be produced by different aggregation functions and/or on the basis of different individual *input* beliefs and desires. When a corporation fires one employee, for instance, these input beliefs and desires marginally change, but that change need not translate into changes at the group level. Since different aggregation functions and inputs to these functions can produce the same group-level results, List and Pettit conclude that it is frequently impossible to perform a summative reduction of group-level features to individual features. In such cases, groups are the bearer of their features as subjects in their own right.

If this shows that groups can function as agents, the question remains whether they can function as epistemically virtuous or vicious agents. In a response to Lahroodi, Todd Jones (2007) answers affirmatively.[62] Functionalists maintain that groups can be organized to realize various cognitive processes (including belief-forming processes). But "[i]f groups can have cognitive processes", Jones writes, then "they can have better and worse working cognitive processes and thus exhibit epistemic virtues" (p. 441). Indeed, once we view groups

---

[61] See De Bruin (2018) for an analysis of aggregating quantitative, financial judgments in the boardroom.

[62] Byerly and Byerly (2016) also appeal to a functionalist analysis when they argue that corporate virtues and vices are multiply realizable: a group can replace one or more of its members without thereby losing its virtues or vices.

as functional kinds, we must conclude that "there are many different ways for groups to have epistemic virtues" (p. 447). And, we might add, epistemic *vices* too. This is because there are many different ways in which groups can be organized to implement cognitive functions. Of the many different conceivable aggregation functions, which take us from individual to group attitudes, that List and Pettit (2011) allow for, some are conducive to the group's epistemic ends, while others obstruct it. The same goes for the decision procedures the group uses to translate these attitudes into action.

The challenge, then, is to identify those organizational structures within which group members combine so as to function as an agent that exhibits a collective epistemic virtue or vice. Despite the work by Jones, this is still largely an open task. In the section that follows, this chapter builds on earlier work by De Bruin (2015) and takes a closer look at the epistemic misconduct disaster at Boeing.[63]

4.3 Case Study: The Boeing 737 Max Disaster

As we have seen in the previous section, organizations can function as epistemically virtuous or vicious agents in many different ways. One reason for this is that organizations can exhibit a wide variety of what Peter French (1979) calls *corporate internal decision structures* (CIDs). CIDs comprise *responsibility flowcharts* that determine the hierarchical relationships between the organization's members, and *corporate decision recognition rules* that determine the mechanisms by which corporate decisions are made. Often anchored in corporate charters, articles of association, and other official documents, these flowcharts and recognition rules assign particular roles to the members of an organization and determine the rights and duties associated with these roles. For example, CIDs determine the conditions under which someone is authorized to speak on behalf of the organization, but also fix internal hierarchies and how beliefs and decisions are aggregated within the organization.

De Bruin (2015) has argued that CIDs are important loci of corporate epistemic virtue or vice. An epistemically virtuous organization, on this view, is structured such that its responsibility flowchart and corporate decision recognition rules together produce a tendency towards epistemically virtuous behavior and against vice. The CID of an open-minded

---

[63] This chapter departs from existing work on corporate virtue that applies virtue theoretical insights to organizational practice, but focuses on moral virtue (see e.g. Gowri, 2007; Moore, 2005, 2015; Sandin, 2007). Earlier virtue epistemological work in the context of business covered *individual* epistemic virtues (De Bruin, 2013).

organization, for instance, will dispose that organization towards taking contrary ideas seriously. For this to work optimally, organizations must satisfy three conditions. First, epistemically virtuous organizations must ensure that group members exhibit the virtues required by their roles within the organization (*virtue-to-function matching*). Secondly, these organizations must encourage the exercise of these virtues by providing a supportive environment (*organizational support for virtue*). And finally, epistemically virtuous organizations must include safeguards against epistemic vice (*organizational remedies against vice*). While epistemically virtuous organizations typically realize all three of these conditions, a failure to realize any one of them can obstruct the epistemic ends of an organization and thus produce epistemically vicious behavior.[64] To see the practical relevance of this approach when it comes to understanding epistemically vicious behavior in collectives, consider the recent case of Boeing.

4.3.1 Background

Boeing is the largest aerospace company in the world, producing commercial and military airplanes as well as rockets, satellites, and security and defense systems. Founded in 1916, it has an impressive track record. Its bombers played a crucial role in deciding World War II; its 747 Jumbo Jet revolutionized the mass tourist industry; and its contributions to space travel include NASA's first probe to circle the moon as well as part of the rockets NASA later used to land astronauts on the moon.

Boeing's most famous accomplishment is, however, the Boeing 737. It was the best-selling jet in aviation history, until two crashes of its newest generation model, the 737 Max, killed 346 people in 2018 and 2019. At the time of writing this chapter, Boeing's chief rival—the European conglomerate Airbus—has overtaken sales of the 737 with its A320, and sector analysts predict that the 737 is unlikely to catch up anytime soon.

Some historical and technical background is important. Boeing launched the first two generations of 737 jets in the 1960s and 1980s, and faced little competition from rival manufacturers until Airbus introduced its A320 in 1987. By the 1990s, it became clear that

---

[64] It is not the ambition of this chapter to provide a conceptual analysis of corporate virtue. As such, this chapter remains agnostic about the necessary and sufficient conditions for corporate virtue. However, the factors it identifies – the presence of *virtue-to-function matching*, *organizational support for virtue*, and *remedies against epistemic vice* – tend to be sufficient for corporate virtue in typical real life cases.

Boeing had a problem on its hands, as many of its long-time clients showed significant interest in Airbus. To remain competitive, Boeing therefore introduced a third generation 737, the 737 Next Generation (NG), with greater fuel capacity, an updated cockpit, and more seats— ten years after Airbus's A320. As it turned out, however, A320 sales far surpassed sales of the 737NG, and in the mid-2000s, analysts therefore believed that Boeing should make a more radical move and design an entirely new aircraft (Thomas, 2006).

Boeing postponed decision-making on whether or not to design a new jet for years, and in 2010 it was again Airbus that made the first move. Airbus decided against developing a new plane, but chose to refit the A320 with more fuel-efficient engines. This practice is known as *re-engining*.

Under the impression that Airbus had misread the markets, Boeing dismissed the viability of re-engining. The head of Boeing's commercial airplanes division thought that Airbus's re-engining was financially unwise, and would lead to "a plane that carriers didn't really want," and so, he thought, Boeing "could wait until the end of the decade to produce a new plane from scratch" (Gelles et al., 2019, para. 13). Boeing's then-CEO James McNerney likewise stated that "the leader in the clubhouse is the all-new aeroplane" (Weitzman, 2011, para. 5).

Boeing was entirely wrong. Oil prices were surging, and carriers did want more fuel-efficient engines, and they wanted them fast. So when one of Boeing's biggest clients, American Airlines, announced in 2011 that it would move part of its business to Airbus, Boeing was forced to reverse course (Odell, 2011). Yet by that point, Boeing's ignoring evidence about consumer preferences had cost it precious time. While Airbus had been successful in re-engining the A320, Boeing's 737 Max suffered, as we saw, two dramatic crashes shortly after entering service in 2017. It is important to understand what happened from a technical perspective.

Unlike its predecessors, the engine of the 737 Max is attached forward on the wings rather than suspended under the wings. This forward engine placement creates particular aerodynamic challenges. With too much power to the engine, the plane's nose may go up, increasing the so-called *angle of attack* (AOA), which is the angle between the wing and the flow of air. A certain AOA is needed to lift the plane, but if a plane exceeds the optimum AOA its lift will suddenly decrease because the air no longer flows smoothly along the wings but becomes turbulent, a condition called *stall*.

Stalling is dangerous. If your paper airplane goes up too steeply, it does not get very far but falls, like a stone, and crashes. Since any aircraft is at risk of stalling, commercial

aircraft have sophisticated stall control systems in place. Although the technical details of the 737 Max's stall-prevention systems are still under investigation, a key component seems to be a software package called the Maneuvering Characteristics Augmentation System (MCAS), which receives information from an AOA sensor close to the jet's nose, and pushes the nose down when the critical AOA is exceeded. Flight data indicate that MCAS received false input from the AOA sensor. It wrongly suggested the plane was about to stall, and therefore automatically pushed the nose down, crashing the plane (House Committee on Transportation and Infrastructure, 2020). AOA sensors are prone to malfunction, exposed as they are to low temperatures and lightning, and they are sometimes installed incorrectly. But Boeing made its stall-prevention system rely on only one AOA sensor in the re-engined 737.

4.3.2 Virtue-to-Function Matching

At the moment of writing this chapter, there seems to be a fair degree of consensus among experts suggesting that the decision to rely on one AOA was a key — and human — error explaining the two crashes. But who took that decision? Some observers have said the 737 Max was "designed by clowns who are in turn supervised by monkeys" (Bushey and Stacey, 2020). This brings us to the first element of corporate virtue: virtue-to-function matching. Any organization has numerous goals. Boeing has the ambition to be the primary choice among pilots. One of its slogans was: "If it's not Boeing, I'm not going." It wants to develop first-rate technology. It wants to maximize profits for its shareholders, and many other things. Achieving these goals involves accomplishing a wide variety of tasks. The design and construction of a wing, for example, requires modeling the aerodynamics of the wing and applying the materials science behind the composites involved in building the wing. It involves the know-how of technicians who assemble the wing, and the expertise of testers who determine whether the wing is safe and functions as intended. It also requires patent lawyers who scrutinize the project for any liabilities. Moreover, before the project even starts, accountants must draft budgets.

In technology-intensive industries, most jobs have substantial epistemic components. Knowledge (information) must be collected, engineered, stored, processed, evaluated, shared with colleagues, communicated to the workforce, and so on. The thought behind virtue-to-function matching as the first component of an epistemically virtuous organization is that these different types of epistemic work are facilitated by different epistemic virtues. The

virtues of curiosity and wonder facilitate new insights through experimentation, engineering, modeling, and reflection. Humility and temperance help evaluate the relevance and reliability of new information. Sharing and communicating information is an exercise in epistemic generosity. And storing information requires attentiveness and care.

No organization can hope to find members that score high on each of these epistemic virtues. Some persons are curious and creative, others more attentive or generous. To reliably achieve the various epistemic ends of an organization, management should therefore ensure that the members of that organization have the epistemic virtues required by their roles and responsibilities within the organization. While this may be difficult to contest conceptually, empirical practice is often complicated.[65]

The most widely studied function in business scholarship is that of the managing director or CEO of a firm. The CEO is the firm's key representative vis-à-vis its owners (shareholders) and other stakeholders, and must have a clear view of the firm's long-term strategy. The CEO is the first and foremost decision maker of the firm and bears the main responsibility for its decisions.

Enter James McNerney, at the helm of Boeing between 2005 and 2015. A day after McNerney announced his departure at Boeing, the prominent aerospace industry analyst Richard Aboulafia (2015) contributed an insightful profile to the respected American business biweekly *Forbes*. Although Aboulafia opens with the observation that McNerney pleased Boeing's shareholders, he reprimands him for leaving a "toxic legacy" (para. 2). Aboulafia details how McNerney's concern for shareholder interests led him to move production to new facilities, and cut pensions and salaries: "Taking away pensions at a time of record sales is a terrible way to motivate workers to go the extra mile" (para. 6).

The question to ask is whether McNerney's epistemic character traits matched his function as a CEO. Aboulafia thinks not. McNerney had no prior experience with aviation when he started at Boeing. His previous job was CEO at 3M (think Post-Its and face masks). But as Aboulafia says: "If a CEO comes from a different industry and doesn't try to learn what makes aviation distinct, he's likely to apply a one-size-fits-all template" (para. 8). And indeed, like many American companies, 3M had faced increasingly intense competition from low-cost countries, which arguably justified the drastic methods McNerney had deployed to ensure the firm's survival. But aviation is a very different industry, with only two major

---

[65] Further, it may not always be easy to link specific tasks to corresponding virtues. This chapter follows Jason Baehr's (2011, p. 21) meticulously argued taxonomy.

global players and precious little competition from outsiders. In such a market, Aboulafia says, "[a]n *experienced* and *motivated* workforce … is the most important asset a company has" (para. 10, emphasis mine).

Several commentators do indeed implicate an *inexperienced* and *unmotivated* workforce in the safety lapses at Boeing's manufacturing plants and its poor handling of the crashes (Gelles, 2020; Kitroeff and Gelles, 2019). Yet it is important to be cautious and avoid suggesting a direct link between suboptimal virtue-to-function matching—McNerney's lack of curiosity about the aviation industry—and the 737 Max disasters. Corporate epistemic virtue requires more than that.

4.3.3 Organizational Support for Virtue

Besides ensuring virtue-to-function matching, an epistemically virtuous organization should strive to create and maintain an environment that is sufficiently conducive to epistemic virtue in which employees should, among others, feel free to ask questions, share knowledge, criticize each other, and investigate things. In such an environment, senior employees will have to pay attention to what juniors say, openly acknowledge the value of their input, and should not be above changing their minds on the basis of this input. In short, epistemically virtuous organizations should have a system of incentives (in the broadest sense of the word) in place to stimulate and support epistemically virtuous behavior.

The earlier claim that the 737 Max was "designed by clowns who are in turn supervised by monkeys" is surely hyperbole. But reports in the media and official investigative findings provide ample evidence that the 737 Max was designed and produced in a decidedly suboptimal epistemic environment. For example, it appears that the commercial pressures at Boeing obstructed the creativity and innovativeness of its engineers. Here are some examples from the congressional hearings and media reports. Engineers were requested to deliver technical drawings at "double the normal pace" (para. 8), and "sloppy blueprints" (para. 29) were delivered by "rushed designers" (Gelles et al., 2019, ibid.). Engineers were forced to make as few changes as possible to the aircraft so as to minimize the need for new pilot training (as this would make the plane less attractive to prospective buyers who would have to pay for the training). This was felt as considerably frustrating their creativity: "there was so much opportunity to make big jumps, but the training differences held us back" (ibid., para. 38). Rather than harnessing the virtues of its engineers, Boeing held them back.

Information sharing was minimal and discouraged throughout design and production processes. When the plane was finally constructed, Boeing was highly reluctant to share information with pilots, who report not understanding particular signals, and finding no relevant explanation in flight manuals (Fallows, 2019). Even prior to the crashes, pilots complained about "[p]oor training and even poorer documentation" (para. 52) and a lack of information about "the highly complex systems that differentiate [the 737 Max] from prior models" (para. 64), with the result that they "lacked the knowledge" (para. 84) required to fly the plane safely (ibid.). In fact, information about MCAS, the software system implicated in both crashes, was missing from the manuals altogether. Boeing reasoned that since MCAS would operate "in the background" (para. 47) pilots would not need to be briefed on it (Gelles et al., 2019).

To be sure, organizational support for epistemic virtue does not require that everyone knows everything. You do not need to understand Linux code to use the operating system responsibly. But in this case we are talking about *pilots*, who were confronted with alarm signals their documentation failed to explain to them *during a flight*.

The House Committee called this a "culture of concealment" (House Committee on Transportation and Infrastructure, 2020, p. 3). This culture also manifests itself in other areas. A key form of organizational support for epistemic virtue is that people can speak freely, without fear of repercussions.[66] Only if employees can be confident that they can talk to superiors without the risk of losing their jobs or being relocated will they speak frankly. Instead of encouraging its employees to report on safety issues, Boeing swept safety issues under the rug and concealed them from regulators. Boeing employees did bring forward various whistleblower complaints to the effect that superiors actively discouraged them from reporting manufacturing errors and other safety violations. But they were discouraged from doing so, and some faced retaliation when they did. Quality managers who noticed that defective parts were installed in planes were told not to worry and removed from projects if they persisted.

Perhaps Boeing's most lamentable decision was making its MCAS flight control system rely on a single AOA sensor, as mentioned earlier. Employees expressed doubts about "whether the system was vulnerable to malfunctioning if a single sensor failed" (para. 6) as early as 2015 (Gelles and Kitroeff, 2019). For reasons that are not entirely obvious, their concern received no uptake, although it seems likely that Boeing underplayed the danger of

---

[66] Fricker (2020) discusses a similarly vicious organizational culture, at the BBC.

an MCAS failure to ease certification procedures (Gates, 2019). While these procedures should have served as a check on the adverse epistemic conditions at Boeing, preliminary investigative findings suggest they failed to act as an effective remedy against Boeing's epistemically vicious tendencies. This is precisely what the last element of organizational epistemic virtue discussed in this chapter aims to avoid.

4.3.4 Organizational Remedies Against Vice

Organizations must offer supportive environments to enable epistemically virtuous individuals. Most of us are no virtue epistemic superheroes, though, and hence organizations must also have remedies in place that mitigate the effects of epistemic vice. These remedies can be implemented at various levels within an organization. An illustrative example of a *macro-level* remedy shows how organizations can protect themselves from adopting one-sided or biased views. An employee or team has invested considerable time and resources in developing a plan for a new product, and presents it to a decision maker within the organization — the "boss." What will the boss do? In many organizations, bosses decide by themselves, and give the project the green light, or not.

Looks good? Not from a virtue epistemological point of view. A procedure like this is asking for corporate narrow-mindedness. At the level of the corporation, only one side of the story is listened to: the story that puts the project in favorable light. The boss could do much better by asking some person or team with no stakes in the project to come up with as many arguments against the project as possible, and then, with the pros and cons in hand, decide.

Organizing opposition or *dissent* is an essential macro-level remedy against various epistemic vices. There is always a risk of being overconfident about a project you are invested in, of rushing to conclusions, or of narrow-mindedly ignoring evidence that suggests a more downbeat view of your plan's prospects. There will always be team members who do not share relevant information as extensively as necessary. People make mistakes, are forgetful, and succumb to sunk-cost fallacies and continue working on a project even after they see that it is not really worth the investment any longer. To mitigate these and other biases, organizations have to develop remedies such as organizing dissent.

One form of dissent that lies at the heart of the aviation industry centers on independent governmental bodies regulating the industry. In the US that mandate falls on the Federal Aviation Administration (FAA), whose setup and underlying rationale resemble the

Securities and Exchange Commission and the Food and Drug Administration. Boeing employees are in close contact with the FAA at all times. There are good reasons for this: designing an airplane is costly, so you do not want to go through the entire design process only to learn that the FAA refuses certification. But the preliminary House Committee report suggests that Boeing and the FAA may have gotten much too close. A central point of concern is to do with *authorized representatives*. These are people employed and paid by Boeing, but tasked to represent the interests of the FAA. Email and WhatsApp conversations show that these representatives nudged the FAA into accepting the view that MCAS, the flight control system, would merely be a "speed trim function" (p. 3, fn. 16) not requiring additional certification and pilot training (House Committee on Transportation and Infrastructure, 2020). The FAA agreed.

Who pays the piper calls the tune? Whenever sharing your knowledge comes at a cost to yourself or your employer, conflicts of interest are likely to arise. Authorized representatives are not alone here, witness elaborate codes of conduct managing conflicts of interests in health care, financial services, accountancy, engineering, and many other professions. If effective remedies are in place that guarantee objectivity and epistemic independence, many of these potential conflicts can be averted, and authorized representatives were indeed shaped into one of Boeing's key devices to organize dissent. They were, that is, a key remedy against epistemic vice.

The House Committee report strongly suggests, however, that this remedy dramatically failed. Not only did authorized representatives misconstrue the flight control system to the FAA, they also failed to inform the FAA of various safety concerns. For instance, they did not warn the FAA that Boeing sold aircraft with inoperative devices meant to detect AOA discrepancies, although that problem was known internally as early as 2015. When Boeing finally set about fixing this fault in its AOA indicator software in 2017, an authorized representative signed off on Boeing's plan to postpone the required software update to 2020, again failing to inform the FAA. And perhaps most damningly, they concealed crucial safety information during the plane's development. One authorized representative questioned the safety of relying on a single AOA sensor in internal communications, but that concern was brushed aside and not reported to the FAA. Moreover, it turns out that several authorized representatives were aware of a Boeing analysis showing that pilots had at most 10 seconds to respond to unusual signs from the flight control system, and that a failure to act accordingly could be "catastrophic" (House Committee on Transportation and Infrastructure, 2020, p. 3). But they never shared this knowledge with the

FAA. At these and other junctures, Boeing's authorized representatives could have counteracted the epistemic misconduct that was generated by the commercial pressures under which Boeing was producing the 737 Max. By failing to do so, they instead let vice run rampant.

4.4 Conclusion

This chapter critically examined extant conceptualizations of collective epistemic vice and virtue, and defended its own, functionalist account. Following this approach, collective vice and virtue are instantiated when groups are organized so as to function as an epistemically virtuous or vicious agent. While the ways in which group agents can enact virtuous or vicious corporate structures no doubt vary, this chapter singled out three elements of such structures: virtuous organizations ensure that group members have the epistemic virtues required by their role within the organization (*virtue-to-function matching*); they provide *organizational support for these virtues*; and they enact *remedies against epistemic vice*.

Correspondingly, organizations can collapse into epistemic vice if they fail to enact a corporate structure that is virtuous in this way. In order to illustrate this, this chapter included a detailed case study of Boeing's epistemic conduct surrounding the crashes of two 737 Max jets in 2018 and 2019. It showed how Boeing's leadership appears to have lacked some of the virtues required of corporate decision makers; how commercial pressures generated an environment that was not conducive to epistemic virtue; and how Boeing's remedies against vice failed to offset these pressures. None of this is to say that the two deadly crashes are entirely to blame on epistemic problems. But collective epistemic vice undoubtedly played a part.

Chapter 5 – Improvised Joint Action

5.1 Introduction

One of the central questions in the field of social ontology concerns the nature of shared agency. What is special about two or more individuals acting *together*? In their analysis of shared agency, social ontologists typically focus on instances of what Michael Bratman (2009, 2014) calls *modest sociality*: small-scale examples that involve up to a handful of individuals acting together in order to realize a shared intention, usually without power asymmetries between them.[67] These theorists ask how we should analyze such jointly intentional acts as painting a house together, performing a pass play in football, or executing a group dance performance. While the dominant voices in social ontology disagree in their analysis of joint phenomena like these, many of them accept as uncontroversial the claim that joint intentional action requires the fact of the shared intention to be *out in the open*, publicly accessible to all participating agents.[68] This common knowledge condition usually entails that two or more individuals cannot engage in joint intentional action unless they know or believe in common that they share the intention or goal of doing something together and that they each intend to do their part in realizing that goal or intention.[69]

In recent years, the common knowledge condition has attracted a fair share of criticism for being too demanding (Blomberg, 2016; Paternotte, 2015). The crux of these critiques is that strong epistemic constraints on joint action problematically rule out certain instances of *minimal* or *weak* joint action (Kutz, 2000; Ludwig, 2007). In cases of weak joint action, two or more individuals manage to act together even though they lack detailed knowledge of each other's participatory intentions. Despite evidence that they are capable of

---

[67] Though *cf.* Salomone-Sehr (2022). Not all theorists agree that sharing an intention is a necessary condition for joint action.

[68] Chant and Ernst (2008) are one of the first to comment on this consensus in a field otherwise characterised by deep philosophical disagreements. Hereafter I will use the terms *joint action*, *joint intentional action*, and *joint (intentional) activity* interchangeably.

[69] A version of this claim is present in the work of e.g. Michael Bratman (1990, 2014), Margaret Gilbert (1989, 2013), Seumas Miller (2001), Philipp Pettit and David Schweikard (2006), and Raimo Tuomela (2007). A notable exception is John Searle (1990).

jointly coordinated action, moreover, the common knowledge condition also appears to rule out the possibility of joint action for some young children and other cognitively underdeveloped agents (Butterfill, 2012; Tollefsen, 2005). This is true specifically of agents who lack a theory of mind and therefore cannot (or least not *yet*) form beliefs about the intentions of others. In this chapter, I identify another type of action that is at odds with the common knowledge condition – namely, intentionally *improvised* joint action. Since so much of our lives is improvised, this gives us a strong reason to scrutinize the role of common knowledge in the dominant theories of shared agency.

In the context of this dissertation, providing a compelling analysis of improvised joint action is especially important because improvisational skills are crucial to many conceptions of virtue (e.g., Nussbaum, 1990). The central role of improvisational skills in theories of virtue derives from what John McDowell (1979) calls the "thesis of uncodifiability" (p. 343). The thesis of uncodifiability states that the world of ethics is too complex to be captured in terms of universal principles and rigid rules that derive from these. That is why virtue ethicists (and indeed epistemologists) maintain that we should cultivate character traits that prepare us to act or cognize well no matter the circumstances. This requires that we learn how to improvise morally good behavior even in situations that are not obviously covered by existing moral codes. Accordingly, it would speak in favor of an account of group virtue if it could show that *groups* – even the small ones discussed in the literature on modest sociality – are capable of improvisation.

The main argument this chapter develops is that the possibility of improvised joint action demonstrates that the common knowledge condition is not a *necessary* condition for joint action. While there are many cases in which common knowledge of shared intentionality facilitates jointly cooperative behavior, I will argue that improvised joint action belongs to the class of *minimal* or *weak* joint action that does not hinge on the presence of such knowledge. My argument proceeds as follows. In section two, I introduce three dominant accounts of joint action, all of which include a common knowledge condition. Since there is some disagreement over how to define *common knowledge*, section three critically discusses two different ways of making the common knowledge condition more precise. Section four argues that improvised joint action does not involve common knowledge of shared intentionality on either of these variations of the common knowledge condition. In section five, I consider where this leaves us. If we drop the common knowledge condition, where does this leave us? My contention is that a *virtue-theoretic* analysis of the knowledge involved in joint action can provide a way forward. Section six concludes.

5.2 Joint Action

What is special about two or more individuals acting *together*? The examples of joint action I am interested in here characteristically take the form of a complex action, individual parts of which are performed by different agents to produce a jointly intended outcome.[70] When Bob and Edith lift a heavy couch together, for instance, Bob's lifting one end of the couch combines with Edith's lifting of the other, enabling both Bob and Edith to realize their shared intention of moving the couch.

Although it is easy to come up with intuitive examples of joint acts, philosophers have not been able to agree on the necessary and sufficient conditions for joint action. As it turns out, our actions often combine with the actions of others to produce outcomes in ways that fall short of fully-fledged joint acts. Sarah Rachel Chant (2007) illustrates this by imagining two shepherds whose actions accidentally combine to render infertile a piece of land that can only sustain one flock of sheep. In Chant's example, each shepherd lets their flock graze the land on alternating days, neither of them with the intention of ruining the land, and neither of them aware that more than one shepherd is using the land. The ruining of the land is something the shepherds do *together*, but not in the sense of togetherness we are interested in when we talk about joint intentional activity. The challenge is to come up with a model of joint action that captures genuine cases of joint intentional activity while ruling out cases like this.

The main contributors to the joint action literature have different intuitions about what distinguishes cases of fully-fledged joint action from cases that fall short of outright joint activity. We will consider roughly three ways of thinking about joint action in what follows. One well-established tradition characterizes the relevant sense of togetherness in terms of *shared intentions*. Michael Bratman (1999, 2014) and Philipp Pettit and David Schweikard (2006), for instance, each argue that the distinctive feature of intentional joint activity lies in the content of the intentions and associated belief-states held by participating agents.[71] A second common approach emphasizes the importance not of shared intentions, but of shared

---

[70] This chapter puts aside the (interesting) question of how group agency manifests itself in large corporate actors like banks, governments, and other institutions. For ease of exposition (and in line with the literature on shared agency), it focuses on small-scale examples of joint action.

[71] The early Raimo Tuomela (1993, 1995) also conceives of joint action in this way. John Searle (1990) is often credited with introducing the view that intentions characterize the different between joint action and individual action, but for Searle the difference lies not in the *content* of these intentions, but our *mode* of intending them (I-intentions vs. we-intentions).Tuomela makes a similar argument in his later work (see, e.g., Tuomela, 2013).

ends – goals the realization of which require two or more individuals to coordinate their activities (Miller, 2001). And finally, both of these approaches can be contrasted with a third influential tradition, which is championed by Margaret Gilbert (1989, 1990, 2009). As we have seen in chapters two and four, Gilbert characterizes the sense of togetherness involved in joint intentional activity primarily in terms of distinctive *normative commitments.*

Although I briefly comment on the differences between these rival traditions below, my main interest concerns an important respect in which they are alike. Each tradition privileges the role of common knowledge in joint action in a way that is at odds with the possibility of *improvised* joint action. In section four, I argue that this gives us a strong reason to reconsider the role of common knowledge in the dominant models of joint action.

5.2.1 Joint Action as Shared Intention

On the first understanding of joint action we will consider, two or more individuals perform a joint action only when they act on a *shared intention*. In a nutshell, this entails that the persons involved must each *intend* to perform a certain action together. To return to our example of Bob and Edith, this view counts their lifting of a couch together as a joint action only if Bob and Edith each act on the shared intention that "*we* lift the couch." This emphasis on shared intentionality also offers a straightforward explanation as to why Chant's sheepherders are *not* engaged in joint action. Since neither shepherd intends that they jointly ruin the land, a crucial component of joint action is missing.

As some authors have noted, however, the concept of a shared intention raises some philosophical problems of its own (Bratman, 1999; Pettit & Schweikard, 2006; Velleman, 1997). In the philosophy of action, it is common to think of intentions as mental states belonging to individual persons, not as states that can be shared among multiple individuals (Baier, 1970, 1997; Stoutland, 1997). Moreover, it is common to think of intentions as tied to individual persons in a very particular way. When I intend to do something, I *settle* something. That is, I make up my mind one way rather than another, which then rationally constrains me to act in ways that are consistent with realizing my intention. If I intend to go swimming at noon, for instance, it would be irrational to schedule a doctor's appointment at the same time. When Bob and Edith each have an intention of the form "*we* intend to lift the couch together," however, it would appear that they are each settling things for the other, who is subsumed under the "*we*" as well. But that is absurd, for how can we make up somebody

else's mind? After all, whether or not Bob's intending that Bob and Edith lift a couch together puts rational pressures on Edith to act consistently with that intention is not something that Bob, strictly speaking, can control. Then how can he rationally intend it?

There are broadly two ways of responding to these concerns. The first response, proposed by Bratman (1999, 2014), is to view shared intentions as interdependent sets of individual intentions. On Bratman's (2014) view, it is not sufficient for joint action that the persons involved *intend* to act together; their intentions must also "interlock" (p. 103) in appropriate ways by referencing the other persons' intentions. This interlocking condition is met when one person's intention to engage in a joint activity is *conditional* on the other persons' intentions (and vice versa). According to Bratman, then, Bob and Edith share the intention that they lift a couch together when Bob's intention that they do so together *depends* on Edith's intention that they do so together, and when Edith's intention likewise depends on Bob's. Given this reciprocity, neither Bob nor Edith is settling things for the other; rather, they are settling on a course of action *together*.[72] Further, this interdependence ensures that both parties are under rational pressure to act in ways consistent with their shared intention, requiring each do their part. Hence, conceiving of shared intentions as interdependent intentions resolves the philosophical tensions inherent in claiming joint ownership over something as personal as an intention.

A second, closely related response trades on the distinction between intending *to* do something and intending *that* something occur (see, e.g., Pettit & Schweikard, 2006, p. 21; Bratman, 1999, ch. 8). As we have just seen, intending *to* do something is to make up one's mind to do that thing, which entails a strong degree of control over whether or not we *will* or *can* do that thing. This is why I cannot rationally intend to win the lottery, as whether or not I win the lottery is something over which I exercise little control. But intending *that* a certain outcome occur does not require that I have control over whether that outcome materializes in the same way. There is no logical contradiction, for instance, in intending *that* my child takes over the family business or attends a prestigious college, though whether or not they *will* do those things is, strictly speaking, up to them. After all, intending that my child acts in these ways does not put rational constraints on *them*, but on *me* (to act in ways conducive to my intention). Hence, if I intend that my child attends a certain college, I am rationally required to take reasonable steps ensuring that my child has that option.

---

[72] This is roughly the solution that Velleman proposes in his (1997) as well.

Similarly, Pettit and Schweikard (2006, p. 23) argue that there is no logical contradiction in a group of individuals' intending *that* they perform a joint action. When Bob and Edith share the intention that *they* lift a couch together, they are not making up *each other*'s mind, but their own. Of course, this puts both of them under rational pressure to act in ways consistent with realizing that intention. Hence, they must each intend *to* do their part of the joint project (and this, of course, is fully within their control). Should Bob intend to do his part, but Edith fail to reciprocate (despite intending that they, together, move the couch), she is not acting as rationality requires, undermining the jointly intentional status of their activity. The sharing of an intention depends on the relevant contribution of everyone involved – it is a *fusing* or *coming together* of intentions. In the scenario that Edith intends to bail on Bob, one person's contributory intention is lacking.

On the view we are considering, however, merely *intending* to do something together is not sufficient for sharing an intention (and, by extension, for joint action). Certain epistemic conditions must be met as well. An important reason for including epistemic considerations is that intentions alone do not fully capture the sense in which two or more persons are the *joint authors* of their shared intentions and jointly intended acts. If two or more individuals are to constitute themselves as the joint author of a certain act or intention, it matters how they view each other. In Bratman's (2014) words, they must view each other as "intentional co-participant[s]" (p. 48) – intentional agents who are capable of contributing their intention to the set of (interdependent) intentions that makes up the shared intention. But for all we have said so far, the persons involved in joint action may not think of each other in this way at all.

As Pettit and Schweikard (2006) observe, the persons involved may each think of themselves as the "only properly intentional agent involved," regarding the others as "zombies, who would automatically, as if under hypnotism, do what was required of them" (p. 22). If *that* is what they think, however, they are not engaged in fully-fledged joint intentional activity at all. Joint intentional activity is jointly intentional precisely because it depends on the fully intended contributions of each. Hence, if one agent leaves open the possibility that their partners are mere automatons incapable of intentional action, it is difficult to see how they could view themselves as participating in a jointly intentional enterprise – and so, *intend* that enterprise. This brings us to the first epistemic constraint on joint action: a *belief* condition.

Since joint action requires that participating agents think of each other as fully intentional partners, both Bratman (2014, p. 103) and Pettit and Schweikard (2006, p. 23) stipulate that the persons involved must not only intend to perform the joint action (and their

part in it), but also believe that the others intend likewise. To illustrate, suppose that Bob and Edith each intend to lift a couch together, but that Bob believes Edith does not intend to do her part. He believes that Edith often fails to keep her promises, say, or that she would rather keep the couch where it is. In that case, Bob no longer sees Edith as a reliable partner, or indeed a partner at all. Hence, he cannot conceive of their enterprise as jointly intentional, since jointly intentional action *requires* the intentional participation of each. There is something incoherent about intending a joint enterprise without believing that one's intention is shared by one's partners.

Further, both Bratman and Pettit and Schweikard argue that the persons involved should intend to perform the joint action partly *because* they believe their partners share the intention to do so. Even if Bob believes that Edith intends to do her part, this is not yet sufficient for sharing an intention. Shared intentions stand or fall with the contribution of everyone involved, and so each participant should direct their individual intention to a certain outcome in the understanding that their individual intention combines in relevant ways with the contributory intentions of others. Thus, Bob must intend to do *his* part at least partly *because* he believes Edith will do her part. They must, as it were, put forth their wills *as one*. It helps to relate this back to our discussion of intentions as devices that settle things for us. On Bratman's view, shared intentions are interdependent intentions because this is how two or more individuals can settle on a course of action together. But how could we view ourselves as together settling (and so, intend to settle) on a course of action if we did not view our intentions as interdependent?

Though this belief condition is an important epistemic constraint on joint action, the consensus in the literature on shared intentionality is that it is not the only one. The second, and for present purposes more important, constraint is what I will call the *common knowledge condition*. The inclusion of this condition is motivated along similar lines as the belief condition we just discussed. Suppose, then, that all of the conditions for sharing an intention discussed thus far are met. Two or more individuals *intend* to perform a joint action together; they each *intend* to do their part; and they each *believe* that the others intend likewise. What is missing? "For all that these clauses stipulate" (p. 23), write Pettit and Schweikard (2006), the persons involved may not believe that *the others* believe everyone intends to do their part. That is, even if they view each other as intentional co-participants in action (and not as *zombies* or *automatons*), they may each suspect that the others do not view them in the same way. But if that is the case, there is "reasonable ground for denying" (ibid.) that the activity in question counts as a fully-fledged joint action.

Not only is it important that the agents involved each believe that the others intend to do their part, they must each believe that the others view *them* as a fully intentional partner as well. If they leave open the possibility that their partners do not view them as intentional co-participants, the status of their activity as jointly intentional is undermined, for the by now familiar reason that they must each conceive of that activity as dependent on the fully intended contribution of all (*cf.* Blomberg, 2016). To rule out scenarios like this, both Bratman and Pettit and Schweikard introduce a *common knowledge condition*. Roughly, this condition holds that it must be *common knowledge* between the persons involved in joint action that they each intend to perform that action, that they each intend to do their part, and that they each believe the others intend so as well. In paradigmatic instances of joint action, Pettit and Schweikard (2006) write, "everything amongst the parties is above board" (p. 23). Bratman (2014) concurs, writing that "the fact of the shared intention will normally be out in the open," or "publicly accessible" to participating agents (p. 57).[73]

Bratman and Pettit and Schweikard thus invoke the common knowledge condition to accomplish a nearly identical argumentative move. They each start with a set of conditions for joint action that are not *quite* sufficient, rendering their accounts vulnerable to counterexamples. To avoid these examples, they argue that we require the common knowledge condition. On the view that emerges, joint action involves acting on a shared intention, where *sharing an intention* involves at least the following conditions. First, the agents involved must intend that they pursue some joint project. Second, they must each intend to do their part. Third, they must each believe that the others intend that they pursue the joint project and do *their* part as well. And fourth, all of this must be common knowledge between them. As I will show in section four, however, including the common knowledge condition makes this account vulnerable to yet *other* types of counterexamples – ones in which two or more individuals *improvise* and manage to act jointly under conditions of uncertainty. But before we continue, I briefly consider the common knowledge condition as it appears in two other theories of shared agency.

---

[73] Although it true that Bratman (2014) partially motivates the common knowledge condition with reference to concerns about coherence and interdependence shared by Pettit and Schweikard (e.g., he writes that it helps "support the interdependence" [p. 88] of shared intentions), Bratman further believes it facilitates one of the hallmark features of sharing an intention, which is to *plan* carrying out that intention. How could we plan our joint action if we do not know in common that we share an intention? I revisit this aspect of Bratman's account in section three.

5.2.2 Joint Action as Collective Ends

If models of shared intentionality were the only models of joint action vulnerable to counterexamples based on the common knowledge condition, that would be bad enough. But unfortunately, close relatives of these models run into similar problems. One of these relatives is Seumas Miller's (2001) teleological account of joint action.

According to Miller, the distinctive feature of joint action is not that it is caused by shared intentions, but that it is directed towards *collective ends*. A collective end is an end shared by multiple individuals – e.g., that a certain state of affairs obtains – which depends for its realization on the actions of more than one of these individuals. On Miller's view, then, joint action is interdependent action towards an end that is held in common. To return to our initial example, suppose that Bob and Edith each have the end of together moving the couch they are sat on to an adjacent room. Then Bob and Edith share the end of moving the couch together, which depends for its realization on the contributory actions of both Bob and Edith.[74] Indeed, neither Bob nor Edith could move the couch *together* without the other, so fulfilling their shared end requires that they each do their part. If Bob and Edith *do* in fact do their part with an eye to realizing their collective end, and so together bring about the state of affairs pursued by both, Bob and Edith satisfy Miller's conditions for joint action, and so perform the joint act of moving the couch into another room.

Although realizing a collective end requires interdependent action, Miller emphasizes that not every instance of interdependent action qualifies as *joint* action. Consider, again, Chant's sheepherder case. As Chant envisions her scenario, no single shepherd has enough of an environmental impact to ruin the land on which their flock grazes. But the actions of two shepherds *together* combine to render that land infertile. In a sense, then, the ruining of the land is constituted by interdependent action. Intuitively, though, it is not *joint* action. Of course, one reason for this is that the shepherds lack the *end* of making the land infertile. But according to Miller, that is not the only reason. What "distinguishes joint action from

---

[74] Crucially, for Miller, collective ends are ends that are necessarily shared. We can imagine a scenario, for instance, in which both Bob and Edith have as their end that the couch they are sat on is moved into another room, but in which they do not care much whether they move it together or by themselves. In that case, though Bob and Edith share the end of moving the couch insofar as they aim at realizing the same state of affairs, their end is not collective because neither Bob nor Edith requires the other in order to realize it. In fact, their end could be realized by a third party moving the couch independently of both. Since nothing about having individual ends like these necessitates interdependent action, Miller claims, they cannot be constitutive of joint action.

interdependent action that is not joint," he writes, is "mutual knowledge" (p. 60). Having (and acting on) a collective end is also subject to *epistemic constraints*.

Miller specifies the details of these epistemic constraints along similar lines as Bratman and Pettit and Schweikard. According to Miller, joint action requires that the collective ends involved are "open" (p. 59) to all participants. He cashes this openness out in terms of "mutual true belief," using the term interchangeably with "mutual knowledge" (ibid.). When two or more individuals have a collective end in the sense required for joint action, then, it will be mutual knowledge between them that they have that end. On Miller's definition, two or more individuals have mutual knowledge of a proposition, *p*, *iff p* is true, they each believe that *p*, they each believe that they each believe that *p*, and so on. So, in Bob and Edith's case, Bob will not only truly believe that Edith has the collective end of moving the couch, and Edith that Bob has that end; they will also truly believe that they each believe this, believe that they each believe that they each believe this, and so on. Miller motivates his mutual knowledge condition in much the same way as Pettit and Schweikard. The label of *joint action* implies *intentional* action, and how could we intentionally do something together if we did not mutually *believe* we were doing something together?

As we will see in section three, Miller's characterization of mutual knowledge and belief roughly corresponds to a classical – *formal* – analysis of common knowledge. In this, Miller appears to be using the term *mutual knowledge* somewhat idiosyncratically. Whereas (formal) epistemologists tend to reserve the concept of mutual knowledge for finite hierarchies of knowledge or belief (see, e.g., Lederman, 2018a; or Vanderschraaf, 2013, for an overview), Miller's use of "and so on" (p. 59) in his definition of mutual knowledge appears to imply an infinite hierarchy of knowledge or belief, which is typically defined in terms of *common knowledge*. Unfortunately, we will see that this renders his account vulnerable to the same kinds of counterexamples as the accounts of Bratman and Pettit and Schweikard. Before I turn to these examples, however, I briefly discuss one last account of joint action: Gilbert's plural subject account.

5.2.3 Joint Action as Joint Commitment

Of the accounts of shared agency under consideration in this chapter, Gilbert's (1989, 1990, 2013) plural subject account is the only non-reductive one. While Bratman, Miller, and Pettit and Schweikard claim that we can fully explain joint action in terms of *individual* intentions

and belief-states, Gilbert posits that we require a new, irreducibly collective concept – that of a *joint commitment*.

As we saw when we considered Fricker's account of group character in chapters two and four, these joint commitments are a constitutive element of what Gilbert calls *plural subjects*. According to Gilbert, group agency is realized when two or more individuals instantiate a plural subject by jointly committing to upholding certain intentions, beliefs, or other features as a body. To take our by now familiar example, when Bob and Edith perform the joint action of lifting a heavy couch, they form a plural subject to the extent that they have jointly committed to the intention of lifting the couch. This joint commitment makes it true that we can ascribe that intention (and the resulting action) to *the group* as a plural subject (*Bob-and-Edith*), regardless of the intentions and contributory actions we ascribe to its individual members (*Bob* and *Edith*).

Crucially, Gilbert believes that joint commitments are not mere sums of individual commitments. We can see this if we attend to the distinctive normative expectations that accompany joint commitments. If I individually commit myself to a certain course of action, then I give myself a strong reason to act accordingly. Other things being equal, however, I am free to rescind that commitment whenever I please. Nothing in the nature of an individual commitment entails that I owe it to anyone but myself to follow through. Joint commitments are different. When two or more individuals *jointly* commit to a certain course of action, they incur certain obligations directed towards each other. Moreover, these obligations are irreducibly collective insofar as they can only be dissolved with the consent of all agents that are bound by them.

Gilbert (1990) famously illustrates this using an example that involves two individuals going for a walk together. What, she asks, makes walking together a *joint* activity? Well, suppose that two individuals, Jack and Sue, *individually* happen to go for the same walk at the same time. If Jack wants to stop walking halfway through and boards a bus, that is perfectly fine; Sue does not have the standing to rebuke him. Contrast this with the scenario in which Jack and Sue agree to go on a walk *together*; that is, the scenario in which they have *jointly committed* to going for a walk. Assume that this scenario is outwardly indistinguishable from the first one: they go for the *same* walk, at the *same* time, and execute the *same* bodily movements. In this scenario, however, it strikes many that Jack does *not* have the standing to rescind his commitment unilaterally. Agreeing to go for a walk with Sue, he *owes* it to her to continue walking. To be released from their joint commitment, Jack requires permission from Sue. Failing this, Sue has the standing to rebuke Jack for violating their agreement when he

boards a bus halfway through their walk. Unlike individual commitments, then, joint commitments mutually bind two or more individuals together in a special kind of normative relationship.

Common knowledge enters into Gilbert's picture of joint action in the *etiology* of joint commitments; that is, in the history leading up to the formation of a joint commitment. Thus, Gilbert (2009) writes:

> How, then, is a joint commitment created? In the basic case, on which I focus here, each of two or more people must openly express his personal readiness jointly with the others to commit them all in a certain way. I mean to imply that each is indeed personally ready for this, and that he expresses this readiness. Once the concordant expressions of all have occurred and are *common knowledge* between the parties, the joint commitment is in place. (p. 180, emphasis mine).

For Gilbert, two or more persons cannot be jointly committed to something unless it is common knowledge between them that they are each personally ready to do so. In typical cases, moreover, this knowledge is in place because each party has *expressed* their personal readiness to jointly commit to something. These expressions can take various forms. For instance, Jack and Sue can jointly commit to their walk by agreeing do so verbally, as when Jack asks "Want to go for a walk?" and Sue answers affirmatively. But not all expressions of personal readiness are verbal. A cashier, for example, may signal their personal readiness to complete store transactions simply in virtue of sitting behind the store counter wearing the appropriate uniform.

Gilbert's common knowledge condition seems natural enough. There is something Kafkaesque about the possibility that two or more persons could be jointly committed to something without commonly *knowing* that they are so committed. Indeed, for Gilbert, jointly committing to something involves holding each other to certain expectations. But how could we coherently hold each other to expectations the other does not know we have of them? As I show in section four, however, Gilbert's common knowledge condition does not square well with some intuitive cases of joint action under conditions of uncertainty. This gives us a strong reason to reconsider the role of common knowledge in her account as well.

5.3 Common Knowledge

As the preceding section shows, the concept of common knowledge does considerable work in the dominant models of joint action. This makes it all the more surprising that joint action theorists tend not to discuss the concept of common knowledge in too much detail. Bratman, for example, explicitly claims to be using an "unanalysed concept" (1999, p. 117) or "intuitive notion" (2014, p. 5) of common knowledge. Pettit and Schweikard (2006) endorse the "usual" (p. 24) hierarchical model of common belief, but do not say much more. And Miller (2001, p. 59, fn. 12) simply points the reader to a diverse literature on the nature of common knowledge and mutual belief. Gilbert (1989, 2008) is a notable exception here, devoting considerable space to developing an account of common knowledge loosely based on David Lewis's (1969) work on convention.

Developing a working conception of common knowledge is, nevertheless, important, because there exist multiple competing analyses of common knowledge in the literature. Moreover, each of these has different implications for models of joint action. Following Harvey Lederman (2018a), this section therefore briefly considers two central conceptions of common knowledge. The first is a *formal* conception of common knowledge, which predominates in economics, game theory, and formal epistemology. The second is a *public information* conception of common knowledge, which aligns more closely with our informal, everyday understanding of common knowledge.[75] In section four, I argue that neither conception of common knowledge can account for all intuitive cases of improvised joint action.

Although we find elements of both conceptions of common knowledge in recent work on joint action, most joint action theorists appear to endorse a *formal* conception of common knowledge. On formal conceptions of common knowledge, two persons, *A* and *B*, *commonly know* a proposition, *p*, if and only if A knows that *p*, B knows that *p*, *A* knows that *B* knows that *p*, *B* knows that *A* knows that *p*, *A* knows that *B* knows that *A* knows that *p*, *B* knows that *A* knows that *B* knows that *p*, and so on. A similar analysis applies to common belief (or "mutual belief" [p. 59] as Miller [2001] appears to be using the term), such that two persons *commonly believe p* if and only if they each believe that *p*, they each believe that they each

---

[75] Though the received view is that these conceptions of common knowledge are logically equivalent, this idea has recently been challenged (see, e.g., Lederman 2018a). At any rate, the public information conception I will be working with in this chapter is based on Gilbert's (1989), which we will see is not equivalent to the formal conception under consideration (Vanderschraaf, 2013).

believe that *p*, and so on. When Pettit and Schweikard (2006) adopt the usual hierarchical model of common belief, it is arguably this formal definition of common belief that they have in mind. Bratman (1999, p. 111; 2014, p. 5) likewise appears to endorse this view when he characterizes common knowledge of shared intentionality as involving nested beliefs about each other's contributory intentions. Miller (2001) follows the above formula almost verbatim, stating that joint action requires that participating agents "mutually truly believe" (p. 59) that they are interdependently acting towards a collective end in this iterative way.

The main benefit of adopting a formal conception of common knowledge is its strong explanatory power. As we have seen, joint action theorists typically invoke the concept of common knowledge to account for the transparent nature of paradigmatic examples of joint action. In paradigmatic examples of joint action, that is, everyone involved is on the same page vis-à-vis their shared intentions, goals, or normative commitments. At any rate, they do not doubt that they are engaged in a joint project. The presence of common knowledge on this formal definition gives us a neat story as to how this could be, since it prevents doubts from arising at infinite levels of belief. If Jack and Sue commonly know that they share the intention of going for a walk together, neither Jack nor Sue has to wonder whether the other knows that they share this intention, whether the other knows that they both know this, and so on. As Gilbert (1989) puts it, common knowledge thus ensures that it is "no surprise" (p. 195) they each intend to do their part. Jack and Sue can therefore safely proceed on the assumption that they both intend to go for a walk together, removing an important obstacle to constituting themselves as the joint author of their shared intention of going for a walk.

In this way, common knowledge also facilitates one of the crucial features of shared intentions that Bratman (1999, 2014) draws our attention to; namely, their role as *plan states*. When you and I share an intention, Bratman points out, it is typically the case that our shared intention guides our plans to realize that intention. Since shared intentions are interdependent, this requires that we coordinate our plans. But the received view in much of contemporary philosophy, economics, and game theory is that we cannot successfully coordinate our plans for jointly intentional activities unless it is common knowledge between us that we share an intention (see, e.g., Paternotte, 2011; Lederman, 2018a, 2018b for an overview). If either of us does not know that the other shares our intention, it would be irrational for us to act on that intention (for what if they do not?). Moreover, our situation would not be much better if all we did know was that the other *does* share our intention. After all, unless we know that the other *knows* that we know this, we cannot be sure that it is rational for *our partner* to coordinate their actions with ours. It should now be relatively easy to see that nothing short of

common knowledge will do for the purposes of coordinating our plans, for no matter how many finite epistemic iterations we add, there is room to doubt that the other has sufficient information to rationally coordinate their plans with ours.[76]

Appealing to formal definitions of common knowledge is, however, controversial, as their explanatory power is often said to be bought at the expense of realism. The most common form of this objection challenges the idea that cognitively finite human beings could ever form infinitely many iterative beliefs (see, e.g., Clark, 1996; Clark and Marshall, 1981; Heal, 1978).[77] If that is correct, then how could a formal concept of common knowledge ever accurately represent our epistemic state?

Despite its intuitive appeal, this demandingness objection is not as powerful as it appears to be. For one, it *does* seem that we are, at least in principle, capable of holding infinitely many beliefs (Lederman, 2018b; Paternotte, 2011). Take any natural number n, and I believe that it is followed by $n + 1$ (even if I have not explicitly formulated this belief for every natural number). But secondly, following Barwise (1988) we should distinguish between the task of *defining* common knowledge and analyzing its real-world applications. The formal definition of common knowledge is a useful idealization, allowing us to model human behavior in highly idealized settings, but we should not expect to find this idealized notion fully instantiated in less than ideally rational agents operating in sub-optimal conditions.[78] That is not to say, however, that less than fully rational agents cannot instantiate approximations of it. Nor does it mean, as Paternotte (2011) observes, that cognitively finite beings like us cannot use the concept of common knowledge in our everyday reasoning.

More worrying for joint action theorists is that formal conceptions of common knowledge make common knowledge highly precarious, rendering models of joint action that depend on it vulnerable to counterexamples. Although common knowledge may not require that we explicitly adopt infinitely many beliefs about each other's belief-states, it *does* require

---

[76] This line of reasoning has been taken to absurd conclusions in e.g. Rubinstein's (1989) *Coordinated Attack* and *Electronic Mail* games, which show that the common knowledge requirement conflicts with the powerful intuition that it can be rational for us to coordinate in epistemic states short of common knowledge. This has led some authors to suggest that we should drop the idea that coordination requires common knowledge (e.g., Lederman 2018b, 2018c), while others suggest that we should adopt a different, less demanding analysis of common knowledge (e.g., Paternotte, 2011).

[77] Though *cf.* Stalnaker (2009), who argues that common knowledge of this sort is, in fact, easier to achieve than hierarchical beliefs with a large but finite level of complexity.

[78] Lederman (2018b) has recently argued that even ideally rational agents may not be capable of forming common knowledge of this idealized kind unless they have direct access to each other's minds.

that we do not possess defeaters for these potential higher-order beliefs. Blomberg (2016, pp. 318-19) exploits this requirement using a case I paraphrase below:

> Hector and Celia are building a block tower together. They each intend that they do so together; they each intend to do their part; they each believe that the other intends to do *their* part; and they each intend to do their own part *because* of believing this. They each take turns putting down blocks. Here's the rub: Hector mistakenly believes that Celia mistakenly believes that he does *not* intend to do his part, taking him to be playing a silly game that consists in covering the top faces of the blocks she puts down (i.e., without the intention of building a tower, but *with* the intention of displaying behavior that eventually leads to a built tower). Since that pattern of behavior is consistent with building a block tower, Hector does not seek to correct the mistaken belief he attributes to Celia. After some time, they successfully complete their block tower.[79]

Blomberg constructs his case such that Hector and Celia meet each of Bratman's and Pettit and Schweikard's conditions for joint action except the common knowledge condition. This is because Hector's false belief that Celia does not take him to have the intention of building a block tower prevents him from forming the higher-order belief (*defeats* the higher-order belief) that Celia knows they are pursuing a joint project.

Since the common knowledge condition fails in Blomberg's case, Pettit and Schweikard cannot count it as an instance of joint action. According to Blomberg, however, there is no principled reason to withhold the status of joint action from Hector and Celia's activity. In fact, he claims, the case exhibits many of the hallmark features of a fully-fledged joint action: Hector and Celia each intend that they build a tower together; their intentions are interdependent in the right way, so they settle on building the tower *jointly*; and they both have sufficient information to rationally coordinate their actions. Why, Blomberg (2016) asks rhetorically, "isn't such a joint performance then an intentional joint action" (p. 319)? It certainly seems that Pettit and Schweikard's worries about zombies are misplaced here. That Hector thinks Celia is mistaken about his intentions does not preclude him from treating her as a fully "intentional co-participant" (Bratman, 2014, p. 44), and it certainly does not license him to think of himself as "the only properly intentional agent involved" (Pettit and

---

[79] Note that the problem Blomberg identifies is not one of *action description*. Building a block tower requires that the tower terminates. The intention Hector falsely believes Celia attributes to him involves a type of behavior that could go on *ad infinitum*.

Schweikard, 2006, p. 22). Hence, Blomberg takes his case to constitute a counterexample to theories of joint action that rely on a common knowledge condition.[80]

At this point, proponents of the common knowledge condition could attempt to strengthen their case by appealing to an alternative analysis of common knowledge, one which Lederman (2018a) refers to as a "public information" or "informal" (p. 128) analysis of common knowledge.[81] While it is often argued that this position coincides with a formal conception of knowledge, Lederman urges us to resist this conclusion. As we shall see shortly, moreover, Gilbert (1989) develops a version this conception of common knowledge that departs from a formal analysis in significant ways.

On a public information conception of common knowledge, a proposition, *p*, is common knowledge between two or more persons if *p* is publicly accessible to all of them. Whether a certain proposition is accessible in this way depends on roughly two conditions. First, public accessibility depends on relevant features of the environment, often called the "set-up" (Heal, 1978, p. 125) or "scene" (Lederman, 2018a, p. 186). In particular, the set-up must be such that it both entails *p* and that everyone involved knows the set-up exists. In a parliamentary session, for instance, the set-up is such that it entails all sorts of propositions about current political debates. Thus, if a politician uses the session to proclaim a certain policy agenda, it *entails* that he or she has that agenda. But the set-up is also *public* insofar as everyone attending or watching the session *knows* they are witnessing a parliamentary session (and so, that they are listening to the same parliamentary speeches, and so on).

Second, public accessibility depends on the standards of reasoning employed by the persons involved (Lewis, 1969). Specifically, these must be such that if *one* person infers from the set-up that *p* (and that everyone else knows the set-up exists), they can *all* infer these things. Combined, this means that the parties involved can each infer that everyone else can infer *p* (and that everyone else can infer that everyone else can infer this, and so on), rendering *p* publicly accessible to each of them. Hence, if the persons attending a parliamentary session share the same standards of reasoning, they can each infer that they can all make similar inferences regarding the policies announced at the session, and so treat these as objects of common knowledge. That is, they can treat these policies as something they are all in a position to know (and *should* know if they follow through on their inferences.)

---

[80] See Ludwig (2007), pp. 387-388 for a similar case (and a similar conclusion).
[81] Philosophers who object to formal definitions as unrealistic often prefer a public information view of common knowledge as well (e.g. Heal, 1978; Paternotte, 2011).

With these conditions in place, it is easy to see the appeal of the widespread view that formal and informal conceptions of common knowledge coincide. Given the assumption that the parties involved are perfectly rational, share the same standards of reasoning, and are capable of forming an infinite number of beliefs, they will deduce an infinite number of iterative beliefs about each other's belief-states (though *cf.* Lederman, 2018a). In her account of common knowledge, however, Gilbert (1989) develops a version of the public information conception of common knowledge that clearly departs from a formal conception. According to Gilbert, a proposition, *p,* is common knowledge between us if it is "open*" (pp. 191-95) to us, where *p* is open* to us if and only if an ideal or "smooth" (p. 189) reasoner in our place would infer both that *p,* and that the others, were they ideal reasoners, would infer that *p* as well (and so on).[82] On Gilbert's proposal, then, common knowledge depends not on the standards of reasoning employed by *actual* persons, but by their ideally rational (that is, logically omnipotent) counterparts. Since this does not require that these actual persons *themselves* make inferences regarding *p,* they can have common knowledge without reasoning through an infinite number of beliefs.

Gilbert's account of common knowledge helps take the sting out of Blomberg's case in the following way. As we have seen, a public information approach does not reduce common knowledge to the mental states of the persons involved, but makes it dependent on features of the environment. Whether Hector and Celia know in common that Hector intends to do his part of building a block tower, then, depends not on the beliefs *actually* held by Hector and Celia. Rather, it depends on the beliefs their ideally rational counterparts are licensed to infer from the situation in which they find themselves. What these beliefs are is not entirely clear from Blomberg's presentation. Indeed, Blomberg (2016) suggests that Hector could be justified in holding his mistaken belief about Celia. "Such false higher-order beliefs," he writes, "are arguably a common upshot of insecurities and mild forms of paranoia that are often present in human relations" (p. 319). But it is not clear that Hector's ideally rational counterpart would suffer these insecurities as well. Indeed, it seems plausible that a logically omnipotent reasoner can see through "mild forms of paranoia," and would infer that, in all likelihood, Celia knows that Hector intends to do his part after all. In that case, however, Hector and Celia *would* have common knowledge on Gilbert's analysis, and so Blomberg's

---

[82] Gilbert construes this openness* primarily in terms of the *perceptions* of the agents involved. On this view, say, a bar fight, is open* between us only if we both perceive the bar fight and perceive each other perceiving the bar fight. As Paternotte (2014) points out, openness can also be grounded in what he calls "typical" events: these are events that we do not perceive each other perceiving, but of which we *know* that a typical person like us has experienced them.

case would not work as a counterexample to all theories of joint action that employ a common knowledge condition.

In the next section, I develop a case that is at odds with the common knowledge condition on both conceptions of common knowledge discussed here. Specifically, I will argue that neither what Lederman (2018a) calls *formal,* nor what he calls *informal* conceptions of common knowledge can account for an important subset of improvised joint action. Since we have seen that the common knowledge condition figures prominently in three of the dominant accounts of joint action, this invites us to consider whether these accounts are better off with alternative epistemic constraints on joint action. In section five, I argue that a virtue theoretic analysis of the knowledge involved in joint performances offers a promising way forward.

5.4 Improvised Joint Action

This section introduces two cases of improvised joint action. The first case involves improvisation in jazz. Although we will see that this case does not violate the common knowledge condition, it is instructive in showing what will be required to construct a case that does. The second case *does* pose a problem for the common knowledge condition. It involves two football (soccer) players successfully improvising a pass play under tricky conditions. We shall see that none of the theories of joint action under consideration has the resources to accommodate this case at present.

First, consider the following paradigmatic example of improvised joint action: the improvised performance of a jazz quartet. In joint jazz improvisation, a group of musicians must "spontaneously coordinate" a musical performance, "anticipating and coordinating their playing behavior without the guide of musical notation" (Walton et al., 2015, p. 2). Since improvised performances are spontaneous, the members of a jazz quartet must collaborate to "construct and negotiate the flow of the performance from moment-to-moment" (ibid.). Hence, improvised performances proceed in a way that is not pre-planned. Making things up as they go along, jazz musicians flexibly respond to each other's musical cues, sometimes taking the performance in unexpected directions. Though each player in a jazz quartet improvises their own contribution, the collaborative nature of a jazz performance makes it a clear instance of *joint action*. In typical cases, then, the musicians involved share both the

collective end and *intention* of playing together, and are *jointly committed* to executing their performance.

Here, then, is a natural thought. Although the members of a jazz quartet typically know in common that they are to give a joint performance, they lack common knowledge of the specific form their performance is going to take. The content of an improvised performance is by its nature uncertain, precluding the musicians in question from forming detailed beliefs about what each musician is going to contribute to the performance. Indeed, since they make their performance up as they go along, the musicians lack advance knowledge of even their *own* part in the performance. Given the inherent unpredictability of jazz improvisation, this is true even if they are ideal reasoners. Hence, an improvised jazz performance lacks the transparent nature the dominant theories of joint action seem to require. As a result, it fails the common knowledge condition and thus constitutes a counterexample to these theories.

In spite of the *prima facie* appeal of this line of reasoning, improvised jazz performances do not constitute a counterexample to extant theories of joint action. In order to see this more clearly, recall that joint actions are complex actions that consist of multiple individual parts. While theories of joint action aim to define the necessary and sufficient conditions pertaining to this complex as a whole, these conditions do not apply to the individual actions that *make up* the whole. Thus, we should distinguish between the jazz quartet's jointly improvised performance and each individual musician's contribution to that performance. These individual contributions must interlock in appropriate ways for the performance to count as an instance of joint action, but they are not *themselves* subject to the common knowledge condition. That is, it need not be common knowledge between the musicians specifically what acts they will perform. Rather, what should be common knowledge is that they are together engaged in a joint project and that they are each ready to do their part, whatever that part consists in.

But now there is no special puzzle in counting an improvised jazz performance as an instance of joint action. Contradictory though it may sound, a jointly improvised jazz performance is not *itself* an improvised activity. Any improvisation that occurs takes place at the *individual* level. While each individual contribution to an improvised performance is spontaneous (and so does not proceed according to a pre-planned script), the *joint performance itself* is fully intended by all participants, and the musicians know this at all times. Moreover, it does not seem wrongheaded to insist that the members of a jazz quartet require common knowledge of their joint performance for the improvisation to be successful,

even if they have to work out the details as they go along. Indeed, imagine a scenario in which one of the musicians falsely believes they will play from a score (or jointly improvise a freestyle hip-hop act instead). The improvised performance would likely fail miserably unless that musician were quickly disabused of their mistaken belief.

In order to challenge the common knowledge condition, then, we need an example in which not just the individual contributions, but the joint action *as a whole* is improvised. The following case describes a scenario that fits the bill:

> ***Pass Play*** – Sally and Jane are playing on the same team in a football match. Sally, a defender, has the ball. Jane, a striker, is near the opponent's goal. Right as Sally is running the ball up the pitch, a freak weather occurrence causes heavy mist to descend upon the field, making it impossible to see more than a few feet ahead. Really, the game should be cancelled – but it is not, leaving Sally wondering how to take advantage of the situation. She forms the intention that she and Jane complete a pass play, hoping that Jane can use the confusion to score a goal. Though Sally cannot see Jane, she kicks the ball roughly in Jane's direction. As the ball rolls out of Sally's line of sight, it rolls into Jane's. Recognizing Sally's signature spin on the ball (and *another* one of Sally's bold plays!), Jane forms the corresponding intention that she and Sally complete a pass play. Jane intercepts the ball, runs towards goal, and scores.

In *Pass Play,* Sally and Jane successfully – and jointly – improvise a pass play, which is a paradigmatic example of joint action. Unlike the jazz case, however, in *Pass Play* the joint act *itself* is improvised. Both Sally and Jane spontaneously react to changing circumstances, executing the pass play without a pre-planned script.

Now, consider what is common knowledge between Sally and Jane. At time t_0, Sally forms the intention that she and Jane complete a pass play. But because of the mist, Jane does not yet know that Sally is about to initiate a pass play, so there is no common knowledge of this intention. Indeed, there is not yet a shared intention. This changes at time t_1, when the ball rolls into Jane's line of sight and she forms the corresponding intention that she and Sally complete a pass play. But since the ball has rolled out of Sally's line of sight (and, for the purposes of this example, out of earshot), Sally cannot know this – she has no way of verifying whether Jane has recognized her intention that they execute a pass play. Moreover, Jane *knows* that Sally cannot see her, and so Jane knows that Sally does not know that she has recognized Sally's intention of completing a pass play. Still, then, there is no common knowledge of the intention to complete a pass play. This does not change even at time t_2,

when Jane completes the pass play. Hence, *Pass Play* appears to be an example of a joint action that fails to meet the common knowledge condition.

Of course, defenders of the common knowledge condition can respond by denying that *Pass Play* is an example of joint action. That is, they can bite the bullet: since *Pass Play* fails a necessary condition for joint action, they can claim it is a case of merely *interdependent action*. Though the actions of Sally and Jane combine to produce an outcome in which Sally transfers the ball to Jane and Jane scores, proponents of the common knowledge condition can deny this outcome was a *jointly intended* one. In this regard, Sally and Jane would then be like Chant's shepherds, whose actions combine to produce an outcome *without* an appropriate fusing of intentions. If Sally and Jane were never on the same page vis-à-vis their pass play, then how could they achieve the level of transparency required to conceive of their performance as one that is truly *joint*?

Arguing that *Pass Play* is not an instance of joint action, however, is far less persuasive than claiming that Chant's sheepherder case is not an instance of joint action. Indeed, pass plays are classic instances of the type of *modest sociality* (Bratman, 2009, 2014) joint action theorists aim to analyze. Like moving a couch or going for a walk together, pass plays patently belong to the class of activities that involve two or more individuals intentionally coordinating their activities towards realizing a shared end. Accordingly, there are strong pre-theoretical reasons for counting *Pass Play* as an instance of joint action. Moreover, *Pass Play* meets all but the epistemic conditions of the joint action theories under consideration. First, Sally and Jane each have the *intention* that they complete a pass play together. Second, these intentions are interdependent insofar as they reference each other. Accordingly, they *do* view themselves as engaged in a joint project. Third, Sally and Jane each intend to do their part in the pass play. And finally, Sally and Jane aim at realizing what Miller (2001) calls a *collective end*: an end that necessarily depends for its realization on the contributions of more than one agent. To repeat Blomberg's rhetorical question – why *isn't* such a performance a joint action?

Furthermore, *Pass Play* differs from Blomberg's case in that it fails the common knowledge condition on *both* conceptions of common knowledge discussed in section three. Since the mist functions as a screen that prevents Sally and Jane from seeing each other, it straightforwardly blocks common knowledge on a formal definition. Given the uncertainty the mist introduces, neither Sally nor Jane is warranted in forming higher-order beliefs about their joint performance. For all Sally knows, Jane may not have recognized her attempt at a pass play; and for all Jane knows, Sally may *believe* that Jane has not recognized her attempt

140

at a pass play. But the misty conditions also ensure that this scenario involves a set-up that blocks common knowledge on a public information conception of common knowledge. Since they cannot see each other, neither Sally nor Jane can infer from the set-up that the other has recognized their intention to complete a pass play. This is true even if Sally and Jane are the logically omnipotent smooth reasoners posited by Gilbert. After all, logical omnipotence does not entail the ability to see through fog.

Although *Pass Play* does not involve common knowledge of *shared intentionality*, however, Sally and Jane may nevertheless meet Gilbert's conditions for plural subjecthood. Despite not seeing each other, that is, Sally and Jane may be *jointly committed* to completing a pass play. To see this, recall that joint commitments arise when it is common knowledge between two or more individuals that they have each expressed a readiness to enter into these commitments. As we saw in section three, this readiness can be communicated in various ways, including non-verbal ones. It is plausible, for instance, that the very fact of playing on the same football team expresses a *standing* readiness to be committed to various football plays, including impromptu pass plays. It is also plausible that this standing readiness is common knowledge among all players. If this is indeed correct, Sally and Jane know in common that they are ready to commit themselves to pass plays regardless of the mist, making them jointly committed to executing the play described in *Pass Play*. This is shown also by the fact that Sally arguably has the standing to rebuke Jane should Jane fail to complete her end of the pass play despite recognizing Sally's intent. Thus, while *Pass Play* succeeds as a counterexample to Bratman's, Miller's, and Pettit and Schweikard's theories of joint action, there is a plausible construal on which it fails as a counterexample to Gilbert's theory.

Although *Pass Play* may meet Gilbert's common knowledge condition in its current formulation, however, it is fairly straightforward to modify *Pass Play* such that it no longer meets this condition. In order to do so, we must alter the set-up such that it can no longer be assumed that the persons involved have expressed a standing readiness to jointly commit themselves to unexpected football plays. Suppose, for instance, that Sally is playing a casual game of football with her friends in the park. Let's further suppose that Jane, with whom she plays on a professional team, is not a part of this game. In the midst of playing, however, Sally spots Jane taking a leisurely stroll in the park, and decides to pull a prank on her. She spontaneously forms the intention that she and Jane complete a pass play, punting the pall towards Jane. A heavy mist descends, and, though surprised, Jane recognizes Sally's signature spin on the ball. She forms the corresponding intention that they complete a pass play,

intercepting the ball shortly thereafter. When the mist lifts, Sally and Jane quickly find each other and congratulate the other for a play well-executed.

In this modified version of *Pass Play*, Sally and Jane again successfully improvise a pass play under conditions of uncertainty. And again, the mist ensures that they lack common knowledge of shared intentions. But in this case, Jane has *not* expressed a standing readiness to be committed to various football plays. After all, Jane visits the park with the intention of taking a stroll. She has no involvement in Sally's game, nor does she signal in any way that she would *like* to become involved in Sally's game. The fact that Sally and Jane are not jointly committed to executing a pass play is shown also by the fact that Sally does not have the standing to rebuke Jane should Jane fail to recognize her attempt at a pass play. Indeed, if Sally blames Jane for not doing her part, Jane can successfully defend herself by pointing out that she never agreed to executing a pass play in the first place – she is just taking a walk![83]

The possibility of improvised joint action, then, does not square well with the common knowledge condition present in any of the theories of joint action under consideration. Hence, the common knowledge condition cannot be a *necessary* condition for joint action. Since so much of our lives is improvised, our theories of joint action should nevertheless be able to accommodate the possibility or improvised joint action. How should we proceed?

5.5 An Alternative Epistemic Constraint on Joint Action?

The failure of the common knowledge condition in some cases of improvised joint action entails that extant theories of joint action cannot account for an important subset of joint activity. In the context of this dissertation, that failure is poignant especially because improvisational skills are central to many conceptions of virtue (e.g., Nussbaum, 1990). On these conceptions of virtue, the virtuous agent is someone who can often dispense with hard and fast moral rules because they can *improvise* morally good behavior even in situations they have never before encountered. If, as this dissertation argues, groups are capable of manifesting virtuous character traits as well, we thus had better show that they have the capacity for improvisation too. This gives us a strong reason to see if we can give an account of joint action that is not reliant on the common knowledge condition. As we will see shortly, a virtue theoretic account of the knowledge involved in joint performances can help us do just that.

---

[83] See Kopec and Miller (2018) for a similarly styled argument against Gilbert's theory of joint action.

Before turning to a virtue theoretic analysis, however, it is worth considering another strategy for dealing with cases like *Pass Play*. We could, that is, simply choose to abandon the common knowledge condition *without* replacing it with an alternative epistemic constraint on joint action. One reason that joint action theorists like Bratman, Gilbert, Miller, and Pettit and Schweikard insist on the importance of common knowledge is that they believe it is an essential aspect of the "cognitive glue" (Bratman, 2014, p. 87) that binds joint action together. But consider what cognitive glue is *already* present in cases like *Pass Play*: shared intentions, collective ends, and beliefs that the other is at least *likely* to share these intentions and ends as well. According to Blomberg (2016) and others who explore the possibility of weak or minimal joint action (e.g., Kutz, 2000; Ludwig, 2007), the presence of these mental states alone is sufficient to hold together instances of joint action. Thus, Kutz (2000) writes that "joint action as such merely requires that there be sufficient overlap among the object of agents' participatory intentions" (p. 31). Similarly, Ludwig (2007) concludes that analyzing "joint intentional action does not require special attention to mutual beliefs, or any robust belief requirement at all" (p. 387).

Furthermore, simply abandoning the common knowledge condition is particularly attractive because the concerns that motivate the common knowledge condition gain no traction in paradigmatic cases of improvised joint action. Consider, for instance, Pettit and Schweikard's (2006) worry that the participants in joint action think of themselves as the "only properly intentional agent involved" (p. 22). While *Pass Play* involves a lot of uncertainty, neither Jane nor Sally is unsure whether the other is a mere zombie *"who would automatically, as if under hypnotism, do what was required of them"* (ibid.). For their pass play to succeed, it seems more appropriate for Sally and Jane to treat each other as *partners*; that is, as autonomous co-participants in action, each capable of making independent judgment calls as necessary. Consider again Sally, kicking the ball into the mist with the intention that she and Jane complete a pass play. Not knowing whether Jane is in a position to receive the ball, it would be presumptuous of Sally to treat Jane as someone who would "automatically" do what was required of her. What if Jane is nowhere near the ball and the ball gets intercepted? If she sees Jane as an excellent soccer player great at thinking on her feet (hence capable of autonomous course corrections), Sally's play is more intelligible.

Another concern that has motivated the common knowledge condition is the role of common knowledge in facilitating *planning* behavior (Bratman, 1999; Bratman, 2014). While it is no doubt true that common knowledge can make planning a shared activity easier, the absence of common knowledge in cases of improvised joint action does not render planning

behavior impossible. In cases like *Pass Play*, we make up plans as we go along, responding to changing conditions as required by circumstance. Sally had not planned for heavy mist to descend upon the pitch, but when it did, she quickly devised a plan to make good use of it. Though Jane was unaware that this plan depended crucially on her contribution, she inferred that it did as soon as the ball entered her line of sight, at which point she formed the corresponding plan to intercept the ball and attempt to score. Plans that come together in this way are a common occurrence in everyday life and do not require that everyone has common knowledge of these plans at all stages of the unfolding action. Indeed, even Bratman (2014) appears to concede as much when he warns against adopting "a caricature of human agents as constantly planning, eschewing spontaneity, and rigidly following through with prior plans" (p. 24).

Spontaneity is an important aspect of human life, so tying joint action *too closely* to common knowledge of rigidly formed plans seems misguided. *Pace* Ludwig (2007), however, this does not necessarily mean that we should abandon "any robust belief requirement" (p. 387) whatsoever. As Paternotte (2014) convincingly argues, a definition of joint action should minimally "explain or justify" why joint action is "generally successful" (p. 114). It is difficult to do this in the absence of any epistemic considerations. Consider, again, Jane and Sally: it would be quite the miracle if they pulled off their pass play absent *any* beliefs about what they were doing. But though they lack common knowledge of sharing an intention, they both possess a wide range of other beliefs that facilitate successful cooperation. In the course of playing on the same team, for instance, Sally and Jane may have acquired a detailed knowledge of each other's skills, abilities, and preferred playing style. Given their experience playing with each other, Sally may *know* that Jane is adept at thinking on her feet, always receptive to bold plays. Similarly, Jane may know Sally as a creative player who regularly initiates these bold plays, taking advantage of any opportunity that presents itself. Background knowledge of this kind facilitates *all sorts* of cooperative behavior that does not depend on common knowledge of shared intentions.

Arguably, then, it is background knowledge not of each other's intentions, but of each other's *action-tendencies* that is required for joint action. Although more research is required to develop this idea further, my contention is that a *virtue-theoretic* approach can help us make this epistemic constraint on joint action more precise. That we frequently succeed in intentionally enacting joint performances in the absence of detailed knowledge of each other's plans and intentions is no mere luck or happenstance, but at least partly function of the skills, traits, and faculties that reliably enable us to predict how others will act under the

circumstances. In the virtue theoretical tradition, these skills, traits, and faculties are often grouped together under the umbrella of *practical virtue* or *practical wisdom*: features that embody a type of intelligence suited to practical action, including cooperative action. The practically wise person is someone who knows how to be effective at *getting things done*. Just like moral virtues, practical virtues are acquired through practice and habituation. The more we perform a certain action or attempt a certain activity, the better we become at these actions and activities – and the better we become at extrapolating from *these* actions and activities to others that resemble it.

These practical virtues, and the practical wisdom that they embody, are precisely what appear to facilitate successful improvisation. When the players in a jazz quartet succeed in spontaneously coordinating their performance, they draw on the skills and musical knowledge acquired over the course of years of practice. It is because of their ample experience and familiarity with playing jazz motifs and improvising together that they are disposed to anticipate each other's playing behaviour, enabling the joint negotiation of the flow of music. The practical intelligence that they display in doing so does not consist in propositional knowledge about each other's intentions, but in a certain *know-how*. Even if they cannot articulate *why* their improvisation is successful, the jazz musicians have the ability to predict the directions in which their partners will take the performance, for instance by attending carefully to subtle musical cues. Similarly, Jane's and Sally's ample experience of playing football together has led to a practical wisdom that enables them to anticipate the state of play and respond flexibly to changing circumstances. Again, this wisdom does not primarily seem to consist in propositional knowledge about each other's intentions, but in the know-how that enables an excellent football player to think on their feet.

As an alternative to the common knowledge condition, then, we could attempt to develop a virtue-theoretic constraint on joint action. Instead of demanding a detailed account of the propositional knowledge involved in joint action, this constraint would invite us to give an account of the practical wisdom and virtues involved in successful joint activities. As we have seen, a virtue theoretic approach would enable us to solve at least *two* problems for joint action theorists. First, it would enable us to accommodate intuitive cases of joint action that do not meet the common knowledge condition as construed by Bratman, Gilbert, Miller, and Pettit and Schweikard – including cases of *improvised* joint action. And second, it would enable us to meet Paternotte's challenge of explaining why joint actions are generally successful in the *absence* of common knowledge of shared ends and intentions.

5.6 Conclusion

While the majority of this dissertation concerns the agency of incorporated groups such as governments, businesses, and other organizations, this chapter zoomed in on cases of what Bratman (2014) calls *modest sociality*. In cases of modest sociality, small groups of individuals successfully perform small-scale joint actions. If the literature on joint action is any indication, most of our small-scale joint activities involve such straightforwardly cooperative acts as moving a couch together, going for a walk together, or performing a dance routine together. These paradigmatic examples of joint action are transparent to all participants involved: the participants share an intention, commonly *know* they share an intention, and carry out that intention according to pre-formed plans. As we have seen in this chapter, however, some joint activities are messier than that. In cases of improvised joint action, two or more individuals successfully act together *without* transparent plans, beliefs, and intentions.

The possibility of improvised joint action is at odds with what I have called the *common knowledge condition*, which is present in all of the dominant models of joint action. Given the prevalence of improvised joint action, I have argued that this gives us a strong reason to reject the common knowledge condition. In its stead, we should look for other epistemic constraints on joint action. While future research is required to do so, my contention is that a *virtue-theoretic* account of the knowledge involved in joint performances is particularly promising.

Conclusion

The main research question this dissertation sought to answer is whether and how group agents are capable of manifesting virtuous or vicious character traits. After a thorough investigation, I am now in a position to answer the *whether*-question with a resounding *yes*. *Some* groups have character, and we can appraise the collective character of group agents along the same virtue ethical or epistemological lines as the character of individuals. The *how*-question, of course, requires a more detailed answer.

In order to show that groups are capable of realizing character traits, this dissertation tangled with challenging questions relating to the metaphysics of groups. The first of these questions asked what it is that we *do* when we ascribe character traits to collectives of various kinds. According to one answer I considered, attributing character to groups and group agents is a mere linguistic shortcut for talking about the character of individual group members. On this view, when we call an investment bank *reckless* or a political movement *deplorable*, we are really calling its *employees* or *followers* reckless or deplorable.

But while we do at times appear to be using talk about group character in this metaphorical way, we have seen that it is unable to account for all instances in which we attribute character to groups. At least sometimes, we truly appear to attribute character traits to groups and group agents *as subjects in their own right*. To give an account of how this is possible, I drew generously on recent insights from the rapidly evolving field of social ontology. Indeed, one of the main areas in which social ontologists have recently made progress concerns the conditions under which individual group members combine to form group *agents* capable of manifesting a wide range of agential features. The task of this dissertation was to study the conditions under which group members combine to form agents capable of manifesting collective *character traits*.

After considering an alternative rooted in Margaret Gilbert's (1989) theory of group agency, I concluded that a *functionalist* account of group character offers a compelling analysis of a wide range of collective character traits. On the new, functionalist account of group character this dissertation develops, groups instantiate collective character traits when their members are organized so to collectively *function* in a characteristic way. Since there are

many ways in which groups can organize themselves to achieve a particular functional state, there are myriad ways in which these collective character traits can be realized. As we saw in chapters two and four, however, the *corporate structure* of organizations often plays a crucial role in bringing about group character.

In conducting the first systematic defense of the idea that we should analyze group character as *functional states*, this dissertation thus offers a thorough account of how groups can realize collective character traits. But the objective of this dissertation was not just to establish that (some) groups have character, but also to establish the second, perhaps more important claim that we can evaluate collective character traits along virtue ethical or epistemological lines. Indeed, this dissertation has defended at length the idea that the notion of group character opens up new avenues for appraising the moral and epistemic conduct of groups and group agents.

If, as I have argued here, groups can manifest character traits as subjects in their own right, this is an important step towards showing that we can evaluate groups in terms of their virtuous or vicious character traits as well. To show that groups are, indeed, capable of instantiating virtuous or vicious character states, this dissertation opened with a thorough exploration of the parallels between individual virtues and vices and their collective counterparts. Thus, we have seen that in the case of individuals, we frequently evaluate each other not just on the basis of how we act (or cognize), but also in terms of the *kind* of persons we are – e.g., on the basis of the virtues and vices we have developed. But we have also seen that we often appraise the conduct of groups in the same way.

While there is a rich literature that studies the conditions under which groups can be held liable for collective acts, however, the literature on group virtues and vices is relatively scarce. This dissertation has sought to address that gap by studying the conditions under which groups of individuals combine to realize collective virtues and vices. Building on the functionalist analysis of group character offered above, I have thus argued that groups manifest the collective equivalent of moral and epistemic virtues and vices when they are organized so as to function in a collectively virtuous or vicious way. The details of this analysis of collective virtue and vice depend both on the virtue or vice in question and the nature of the group involved. In order to function in a collectively *courageous* way, for instance, a group will need to realize a different organizational structure than required for functioning in a collectively *generous* ways. And that structure will likely look different for a corporation aimed at realizing profits than for a political party aimed at furthering the public good.

Although groups and group agents can instantiate collective virtues and vices, we have also seen that instantiating collective virtues and vices is not always straightforward. Specifically, we have seen that groups are vulnerable to the situationist challenge in much the same way as individuals are. Just like *individual* character traits can be destabilized by situational influences, that is, so can their collective counterparts. Fortunately, these situational liabilities need not concern advocates of a virtue theoretic approach to group character too much. For one, the same conceptual and hybrid defenses virtue theorists have used to defend the concepts of individual virtue and vice can be used to defend the concepts of *collective* virtue and vice. For another, the situationist critique seems less damaging in the case of group agents, since group agents are often designed with respect to achieving a narrowly circumscribed range of objectives. Hence, it seems reasonable to suppose that group agents need not manifest virtue across a wide range of situations that do not relate to these objectives.

As we have seen over the course of this dissertation, Anscombe's call to reintroduce the concept of virtue into the mainstream of philosophy has been answered in more ways than one. Not only have virtue theoretical insights reinvigorated the philosophically central disciplines of ethics and epistemology, I have also shown that they also promise to deliver a deeper understanding of the ethics and epistemology of *groups*. While the account of collective character, virtue, and vice defended in this dissertation has developed a number of ways in which virtue theoretical insights enhance our understanding of group life, it also opens up promising avenues for future research. I sketch three of these below.

On a *theoretical* level, the functionalist account of group virtue and vice developed here opens up research into the *specific* virtues and vices realizable at the collective level. Since different virtues and vices are characterized by different motives, ends, and dispositions, future research can look into the specific functional profiles groups should strive to realize in order to instantiate particular virtues or vices. What, specifically, is required of a *courageous* institution? And under what conditions does a political party qualify as *open-minded*?

On an *empirical* level, the functionalist account developed here can guide research into the situational pressures and influences that aid or destabilize collective character dispositions. While theoretical research can tell us how groups realize collective virtues and vices *in principle*, it is silent on the *actual* challenges groups face when it comes to virtuous or vicious functioning. Bringing these challenges into focus requires expanding the situationist challenge to collectives. When do situational influences disrupt e.g. an

organization's disposition to justice or diligence? And what can organizations do to arm themselves against these influences?

On an *applied* level, a functionalist account of group character can deepen our understanding of how collective virtues and vices have played a role in concrete cases. Just like this dissertation studied the role of collective epistemic vices in the case of the Boeing 737 Max disasters, we can study the role of collective virtues and vices in other problem cases deserving of scrutiny. In the Netherlands, for instance, the tax authorities have recently become embroiled in a widely publicized childcare benefits scandal, with tax officials wrongly accusing thousands of parents of making fraudulent benefits claims (see, e.g., Erdbrink, 2021). What, if any, is the role of (epistemic) vice in institutional failings such as these?

In other words, the work started in this dissertation far from exhausts the possibilities offered by a functionalist account of group character. While we now have a solid foundation for the study of collective virtue and vice, a great variety of research opportunities remain. My work, then, is only just beginning.

Bibliography

Aboulafia, R. (2015, June 24). Boeing will pay high price for McNerney's mistake of treating aviation like it was any other industry. *Forbes*. https://www.forbes.com/sites/richardaboulafia/2015/06/24/boeing-mcnerney-and-the-high-price-of-treating-aircraft-like-it-was-any-other-industry/

Alfano, M. (2012). Expanding the situationist challenge to responsibilist virtue epistemology. *The Philosophical Quarterly, 62*(247), 223-249. https://doi.org/10.1111/j.1467-9213.2011.00016.x

Alfano, M. (2013). *Character as Moral Fiction*. Cambridge University Press. https://doi.org/10.1017/CBO9781139208536

Alfano, M. (2014). Expanding the situationist challenge to reliabilism about inference. In A. Fairweather (Ed.), *Virtue epistemology naturalized: Bridges between virtue epistemology and philosophy of science* (pp. 103-122). Springer International. https://doi.org/10.1007/978-3-319-04672-3_7

Alfano, M., De Ridder, J., & Klein, C. (Eds.). (Forthcoming*). Social Virtue Epistemology*. Routledge.

Alpert, J. I., & Alpert, M. I. (1989). Background music as an influence in consumer mood and advertising responses. In T. K. Srull (Ed.), *NA - Advances in Consumer Research* (Vol. 16, pp. 458-491). Association for Consumer Research.

Alzola, M. (2008). Character and environment: The status of virtues in organizations. *Journal of Business Ethics, 78*(3), 343-357. https://doi.org/10.1007/s10551-006-9335-7

Alzola, M. (2012). The empirics of virtue theory: What can psychology tell us about moral character? In C. Luetge (Ed.), *Handbook of philosophical foundations of business ethics* (pp. 89-107). Springer. https://doi.org/10.1007/978-94-007-1494-6_48

Annas, J. (2003). Virtue ethics and social psychology. *A Priori, 2*, 20-34.

Annas, J. (2005). Comments on John Doris's *Lack of Character. Philosophy and Phenomenological Research, 71*(3), 636-642. https://doi.org/10.1111/j.1933-1592.2005.tb00476.x

Bibliography

Annas, J. (2010). The structure of virtue. In M. DePaul & L. Zagzebski (Eds.), *Intellectual virtue: Perspectives from ethics and epistemology*. Oxford University Press. https://doi.org/ 10.1093/acprof:oso/9780199252732.001.0001

Annas, J. (2011). *Intelligent virtue*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199228782.001.0001

Anscombe, G. E. M. (1958). Modern moral philosophy. *Philosophy, 33*(124), 1-19. https://doi.org/10.1017/S0031819100037943

Appiah, K. A. (2008). *Experiments in ethics*. Harvard University Press. https://doi.org/10.2307/j.ctv19m61c9

Ariely, D. (2008). *Predictably irrational*. Harper Collins.

Aristotle. (2020). *The Nicomachean ethics* (W. D. Ross, Trans.). Oxford University Press. (Original work published 350 B.C.E.). https://doi.org/10.1093/actrade/9780199213610.book.1

Arpaly, N. (2003). *Unprincipled virtue: An inquiry into moral agency*. Oxford University Press. https://doi.org/10.1093/0195152042.001.0001

Asch, S. E. (1955). Opinions and social pressure. *Scientific American, 193*(5), 31-35. https://doi.org/10.1038/scientificamerican1155-31

Ashburn-Nardo, L., Blanchar, J. C., Petersson, J., Morris, K. A., & Goodwin, S. A. (2014). Do you say something when it's your boss? The role of perpetrator power in prejudice confrontation. *Journal of Social Issues, 70*(4), 615-636. doi:https://doi.org/10.1111/josi.12082

Athanassoulis, N. (2000). A response to Harman: Virtue ethics and character traits. *Proceedings of the Aristotelian Society, 100*(1), 215-221. https://doi.org/10.1111/j.0066-7372.2003.00012.x

Athanassoulis, N. (2002). Review of *Morals from Motives*. *Notre Dame Philosophical Reviews.* https://ndpr.nd.edu/reviews/morals-from-motives/

Athanassoulis, N. (2012). *Virtue ethics*. Continuum.

Baehr, J. (2010). Epistemic malevolence. *Metaphilosophy, 41*(1-2), 189-213. https://doi.org/10.1111/j.1467-9973.2009.01623.x

Baehr, J. (2011). *The inquiring mind: On intellectual virtues and virtue epistemology*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199604074.001.0001

Baier, A. (1970). Act and intent. *The Journal of Philosophy 67*, 648-58. https://doi.org/10.2307/2024585

Baier, A. (1997). Doing things with others: The mental commons. In L. Alanen, S. Heinämaa, & T. Wallgren, (Eds.), *Commonality and particularity in ethics*. Palgrave Macmillan. https://doi.org/10.1007/978-1-349-25602-0_2

Balmer, N. J., Nevill, A. M., & Lane, A. M. (2005). Do judges enhance home advantage in European championship boxing? *Journal of Sports Sciences, 23*(4), 409-416. https://doi.org/10.1080/02640410400021583

Balmer, N. J., Nevill, A. M., & Williams, A. M. (2001). Home advantage in the winter olympics (1908-1998). *Journal of Sports Sciences, 19*(2), 129-139. https://doi.org/10.1080/026404101300036334

Balmer, N. J., Nevill, A. M., & Williams, A. M. (2003). Modelling home advantage in the summer olympic games. *Journal of Sports Sciences, 21*(6), 469-478. https://doi.org/10.1080/0264041031000101890

Baron, R. A., & Thomley, J. (1994). A whiff of eality: Positive affect as a potential mediator of the effects of pleasant fragrances on task performance and helping. *Environment and Behavior, 26*(6), 766-784. https://doi.org/10.1177/0013916594266003

Barwise, J. (1988). Three views of common knowledge. In M. Y. Vardi (Ed.), *Proceedings of the 2nd conference on theoretical aspects of reasoning about knowledge* (pp. 365-379). Morgan Kaufmann Publishers. https://doi.org/10.1007/978-3-319-20451-2_37

Bates, T. & Kleingeld, P. (2018). Virtue, vice, and situationism. In N. E. Snow (Ed.), *The Oxford handbook of virtue* (pp. 524-545). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199385195.013.25

Battaly, H. (2010). Introduction: Virtue and vice. *Metaphilosophy, 41*(1-2), 1-21. https://doi.org/10.1111/j.1467-9973.2009.01633.x

Battaly, H. (2014). Varieties of epistemic vice. In J. Matheson & R. Vitz (Eds.), *The ethics of belief*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199686520.001.0001

Battaly, H. (2015). A pluralist theory of virtue. In M. Alfano (Ed.), *Current controversies in virtue theory* (pp. 7-21). Routledge. https://doi.org/10.4324/9781315728001-1

Battaly, H. (2016). Epistemic virtue and vice: Reliabilism, responsiblism, and personalism. In *Moral and intellectual virtues in Western and Chinese philosophy: The turn toward virtue* (pp. 99-120). Routledge. https://doi.org/10.4324/9781315683935-15

Battaly, H. (2019). Vice epistemology has a responsibility problem. *Philosophical Issues, 29*(1), 24-36. https://doi.org/10.1111/phis.12138

Bazargan-Forward, S., & Tollefsen, D. P. (Eds.). (2020). *The Routledge handbook of collective responsibility*: Routledge. https://doi.org/10.4324/9781315107608

BBC News. (2018). Oxfam Haiti allegations: How the scandal unfolded. https://www.bbc.com/news/uk-43112200

BBC News. (2021). Oxfam: UK halts funding over new sexual exploitation claims. https://www.bbc.com/news/health-56670162

Becker, M. C. (2004). Organizational routines: a review of the literature. *Industrial and Corporate Change, 13*(4), 643-678. https://doi.org/10.1093/icc/dth026

Becker, M. C. (Ed.) (2008). *Handbook of organizational routines*: Edward Elgar Publishing. https://doi.org/10.4337/9781848442702

Begby, E. (2013). The epistemology of prejudice. *Thought: A Journal of Philosophy, 2*(1), 90-99. https://doi.org/10.1002/tht3.71

Beggs, D. (2003). The idea of group moral virtue. *Journal of Social Philosophy, 34*(3), 457-474. https://doi.org/10.1111/1467-9833.00194

Bigelow, J., & Pargetter, R. (2007). Integrity and autonomy. *American Philosophical Quarterly, 44*(1), 39-49.

Bird, A. (1998). Dispositions and antidotes. *The Philosophical Quarterly, 48*(191), 227-234. https://doi.org/10.1111/1467-9213.00098

Björnsson, G., & Hess, K. (2017). Corporate crocodile tears? On the reactive attitudes of corporate agents. *Philosophy and Phenomenological Research, 94*(2), 273-298. https://doi.org/10.1111/phpr.12260

Blackburn, S. (2005, 30 September). Simply wrong. *Times Literary Supplement,* 11-12.

Blomberg, O. (2016). Common knowledge and reductionism about shared agency. *Australasian Journal of Philosophy 94*(2), 315-326. https://doi.org/10.1080/00048402.2015.1055581

Blum, B. (2018, June 7). The lifespan of a lie. The most famous psychology study of all time was a sham. Why can't we escape the Stanford prison experiment? *Medium*. https://gen.medium.com/the-lifespan-of-a-lie-d869212b1f62

Boudreaux, C. J., Sanders, S. D., & Walia, B. (2017). A natural experiment to determine the crowd effect upon home court advantage. *Journal of Sports Economics, 18*(7), 737-749. https://doi.org/10.1177/1527002515595842

Bovens, L. (2016). The ethics of Dieselgate. *Midwest Studies In Philosophy, 40*(1), 262-283. https://doi.org/10.1111/misp.12060

Brady, M. S. (2016). Group emotion and group understanding. In M. S. Brady & M. Fricker (Eds.), *The epistemic life of groups* (pp. 95-110). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198759645.003.0006

Brady, M. S. (2018). The role of emotion in intellectual virtue. In H. Battaly (Ed.), *The Routledge handbook of virtue epistemology* (pp. 47-57). Routledge. https://doi.org/10.4324/9781315712550-5

Bratman, M. E. (1999). *Faces of intention*. Cambridge University Press. https://doi.org/10.1017/CBO9780511625190

Bratman, M. E. (2009). Modest sociality and the distinctiveness of intention. *Philosophical Studies, 144*(1), 149 − 165. https://doi.org/10.1007/s11098-009-9375-9

Bratman, M. E. (2014). *Shared agency: A planning theory of acting together*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199897933.001.0001

Broadie, S. (1991). *Ethics with Aristotle*. Oxford University Press. https://doi.org/10.1093/0195085604.001.0001

Broadie, S. (2012). Eudaimonism. In S. Hornblower, A. Spawforth, & E. Eidinow (Eds.), *Oxford classical dictionary* (4th ed.). Oxford University Press. https://doi.org/10.1093/acref/9780199545568.001.0001

Brown, J. (2021). Group motivation. *Noûs*. https://doi.org/10.1111/nous.12366

Brownstein, M., & Saul, J. (Eds.). (2016a). *Implicit bias and philosophy vol. 1: Metaphysics and epistemology*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198713241.001.0001

Brownstein, M., & Saul, J. (Eds.). (2016b). *Implicit bias and philosophy vol. 2: Moral responsibility, structural injustice, and ethics*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198766179.001.0001

Buchanan, D. A., & Huczynski, A. (2017). *Organizational behaviour* (9th ed.). Pearson.

Burgerlijk Wetboek Boek 2. (2017, July 1). Overheid.nl. https://wetten.overheid.nl/BWBR0003045/2022-01-01

Burn, S. M. (2009). A situational model of sexual assault prevention through bystander intervention. *Sex Roles: A Journal of Research, 60*(11-12), 779-792. https://doi.org/10.1007/s11199-008-9581-5

Bushey, C., & Stacey, K. (2020, January 10). Boeing workers mocked regulators over 737 Max approval. *The Financial Times*. https://www.ft.com/content/7785de46-3350-11ea-9703-eea0cae3f0de

Bibliography

Butterfill, S. (2012). Joint action and development. *The Philosophical Quarterly, 62*(246), 23-47. https://doi.org/10.1111/j.1467-9213.2011.00005.x

Byerly, T. R., & Byerly, M. (2016). Collective virtue. *Journal of Value Inquiry, 50*, 33-50. https://doi.org/10.1007/s10790-015-9484-y

Cassam, Q. (2016). Vice epistemology. *The Monist, 99*(2), 159-180. https://doi.org/10.1093/monist/onv034

Cassam, Q. (2019). *Vices of the mind: From the intellectual to the political*. Oxford University Press. https://doi.org/10.1093/oso/9780198826903.001.0001

Chant, S.R. (2007). Unintentional collective action. *Philosophical Explorations, 10*(3), 245-256. https://doi.org/10.1080/13869790701535246

Chant, S. R., & Ernst, Z. (2008). Epistemic conditions for collective action. *Mind, 117*(467), 549-573. https://doi.org/10.1093/mind/fzn033

Choi, S. (2011). What is a dispositional masker? *Mind, 120*(480), 1159-1171. https://doi.org/10.1093/mind/fzr083

Chozick, A. (2016, September 10). Hillary Clinton calls many Trump backers 'deplorables,' and G.O.P. pounces. *The New York Times*. https://www.nytimes.com/2016/09/11/us/politics/hillary-clinton-basket-of-deplorables.html

Clark, H. (1996). *Using language*. Cambridge University Press. https://doi.org/10.1017/CBO9780511620539

Clark, H., & Marshall, C. (1981). Definite reference and mutual knowledge. In A. Joshi, B. Webber, & I. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). Cambridge University Press.

Coakley, M. (2017). *Motivation ethics*. Bloomsbury. https://doi.org/10.5040/9781350004610

Code, L. (1984). Toward a 'responsibilist' epistemology. *Philosophy and Phenomenological Research, 45*(1), 29-50. https://doi.org/10.2307/2107325

Code, L. (1987). *Epistemic responsibility*. University Press of New England. https://doi.org/10.2307/2185214

Cohen, T., & Morse, L. (2014). Moral character: What it is and what it does. *Research in Organizational Behavior, 34*, 43-61. https://doi.org/10.2139/ssrn.2443685

Cordell, S. (2017). Group virtues: No great leap forward with collectivism. *Res Publica, 23*(1), 43-59. https://doi.org/10.1007/s11158-015-9317-7

Courneya, K. S., & Carron, A. V. (1991). Effects of travel and length of home stand/road trip on the home advantage. *Journal of Sport & Exercise Psychology, 13*(1), 42-49. https://doi.org/10.1123/jsep.13.1.42

Crerar, C. (2018). Motivational approaches to intellectual vice. *Australasian Journal of Philosophy, 96*(4), 753-766. https://doi.org/10.1080/00048402.2017.1394334

Crisp, R. (2004). Does modern moral philosophy rest on a mistake? *Royal Institute of Philosophy Supplement, 54*, 75-93. https://doi.org/10.1017/S1358246100008456

Crisp, R. (2010). Virtue ethics and virtue epistemology. *Metaphilosophy, 41*(1-2), 22-40. https://doi.org/10.1111/j.1467-9973.2009.01621.x

Cunningham, M. R. (1979). Weather, mood, and helping behavior: Quasi experiments with the sunshine samaritan. *Journal of Personality and Social Psychology, 37*(11), 1947-1956. https://doi.org/10.1037/0022-3514.37.11.1947

Darley, J. M., & Batson, C. D. (1973). "From Jerusalem to Jericho": A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology, 27*(1), 100-108. https://doi.org/10.1037/h0034449

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology, 8*(4, Pt.1), 377-383. https://doi.org/10.1037/h0025589

de Bruin, B. (2013). Epistemic virtues in business. *Journal of Business Ethics, 113*(4), 583-595. https://doi.org/10.1007/s10551-013-1677-3

de Bruin, B. (2015). *Ethics and the global financial crisis: Why incompetence is worse than greed*. Cambridge University Press. https://doi.org/10.1017/cbo9781139237093

de Bruin, B. (2018). Moral responsibility for large-scale events: The difference between climate change and economic crises. *Midwest Studies In Philosophy, 42*(1), 191-212. https://doi.org/10.1111/misp.12090

de Bruin, B. (2020). Epistemic corporate culture: Knowledge, common knowledge, and professional oaths. *Seattle Law Review, 43*(2), 807-839.

Doris, J. M. (1998). Persons, situations, and virtue ethics. *Noûs, 32*(4), 504-530. https://doi.org/10.1111/0029-4624.00136

Doris, J. M. (2002). *Lack of character: Personality and moral behavior*. Cambridge University Press. https://doi.org/10.1017/CBO9781139878364

Dovidio, J. F., Gaertner, S. L., & Pearson, A. R. (2016). Racism among the well intentioned: Bias without awareness. In A. G. Miller (Ed.), *The social psychology of good and evil* (2nd ed., pp. 95-118). Guilford Press.

Bibliography

Driver, J. (1999). Modesty and ignorance. *Ethics, 109*(4), 827-834.
https://doi.org/10.1086/233947

Driver, J. (2001). *Uneasy virtue*. Cambridge University Press.
https://doi.org/10.1017/cbo9780511498770

Driver, J. (2016). Minimal virtue. *The Monist, 99*(2), 97-111.
https://doi.org/10.1093/monist/onv032

Erdbrink, T. (2021, Jan. 15). Government in Netherlands resigns after benefits scandal. *The New York Times*. https://www.nytimes.com/2021/01/15/world/europe/dutch-government-resignation-rutte-netherlands.html

Fallows, J. (2019, March 13). Here's what was on the record about problems with the 737 Max. *The Atlantic*. https://www.theatlantic.com/notes/2019/03/heres-what-was-on-the-record-about-problems-with-the-737-max/584791/

Feinberg, J. (1968). Collective responsibility. *Journal of Philosophy, 65*(21), 674-688.
https://doi.org/10.2307/2024543

Feldman, M. (2000). Organizational routines as a source of continuous change. *Organization Science, 11*, 611-629. https://doi.org/10.1287/orsc.11.6.611.12529

Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly, 48*(1), 94-118.
https://doi.org/10.2307/3556620

Fischer, P., Krueger, J. I., Greitemeyer, T., Vogrincic, C., Kastenmüller, A., Frey, D., … Kainbacher, M. (2011). The bystander-effect: A meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin, 137*(4), 517-537. https://doi.org/10.1037/a0023304

Flanagan, O. (1993). *Varieties of moral personality: Ethics and psychological realism*.
Harvard University Press. https://doi.org/10.2307/j.ctv1pncrz2

Foot, P. (1978). *Virtues and vices and other essays in moral philosophy*. Oxford University Press.

Foot, P. (2001). *Natural goodness*. Oxford University Press.
https://doi.org/10.1093/0198235089.001.0001

French, P. A. (1979). The Corporation as a moral person. *American Philosophical Quarterly, 16*, 207-215.

French, P. A. (1995). *Corporate ethics*. Wadsworth Publishing.

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198237907.001.0001

Fricker, M. (2010). Can there be institutional virtues? In T. H. Gendler, & J. Hawthorne (Ed.), *Oxford studies in epistemology* (Vol. 3, pp. 235-252). Oxford University Press.

Fricker, M. (2013). Epistemic justice as a condition of political freedom? *Synthese, 190*(7), 1317-1332. https://doi.org/10.1007/s11229-012-0227-3

Fricker, M. (2015). Epistemic contribution as a central human capability. In G. Hull (Ed.), *The equal society: Essays on equality in theory and practice* (pp. 73-89). Lexington Books.

Fricker, M. (2016). What's the point of blame? A paradigm based explanation. *Noûs, 50*(1), 165-183. https://doi.org/10.1111/nous.12067

Fricker, M. (2020). Institutional epistemic vices: The case of inferential inertia. In I. J. Kidd, H. Battaly, & Q. Cassam (Eds.), *Vice epistemology* (pp. 89-107). Routledge. https://doi.org/10.4324/9781315146058-8

Fried, R., & Berkowitz, L. (1979). Music hath charms … and can influence helpfulness. *Journal of Applied Social Psychology, 9*(3), 199-208. https://doi.org/10.1111/j.1559-1816.1979.tb02706.x

Gates, D. (2019, March 17). Flawed analysis, failed oversight: How Boeing, FAA certified the suspect 737 MAX flight control system. *Seattle Times*. https://www.seattletimes.com/business/boeing-aerospace/failed-certification-faa-missed-safety-issues-in-the-737-max-system-implicated-in-the-lion-air-crash/

Gelles, D. (2020, January 10). 'I honestly don't trust many people at Boeing': A broken culture exposed. *The New York Times*. https://www.nytimes.com/2020/01/10/business/boeing-737-employees-messages.html

Gelles, D., & Kitroeff, N. (2019, October 30). Documents show safety concerns at Boeing before deadly crashes. *The New York Times*. https://www.nytimes.com/2019/10/30/business/boeing-muilenburg-testimony-congress.html

Gelles, D., Kitroeff, N., Nicas, J., & Ruiz, R. R. (2019, March 23). Boeing was 'go, go, go' to beat Airbus with the 737 Max. *The New York Times*. https://www.nytimes.com/2019/03/23/business/boeing-737-max-crash.html

Gilbert, M. (1989). *On social facts*. Routledge. https://doi.org/10.1515/9780691214627

Gilbert, M. (1990). Walking together: A paradigmatic social phenomenon. *Midwest Studies in Philosophy 15*(1), 1-14. https://doi.org/10.1111/j.1475-4975.1990.tb00202.x

Gilbert, M. (2004). Collective epistemology. *Episteme, 1*(2), 95-107. https://doi.org/10.3366/epi.2004.1.2.95

Bibliography

Gilbert, M. (2006). *A theory of political obligation: Membership, commitment, and the bonds of society*. Oxford University Press. https://doi.org/10.1093/0199274959.001.0001

Gilbert, M. (2008). Social convention revisited. *Topoi*, *27*, 5-16. https://doi.org/10.1007/s11245-008-9032-5

Gilbert, M. (2009). Shared intention and personal intentions. *Philosophical Studies*, *144*(1): 167-187. https://doi.org/10.1007/s11098-009-9372-z

Gilbert, M. (2013). *Joint commitment: How we make the social world*. Oxford University Press. https://doi.org/10.1515/jso-2014-0032

Gowri, A. (2007). On corporate virtue. *Journal of Business Ethics, 70*(4), 391-400. https://doi.org/10.1007/s10551-006-9117-2

Greco, J. (2010). *Achieving knowledge: A virtue-theoretic account of epistemic normativity*. Cambridge University Press. https://doi.org/10.1017/cbo9780511844645

Harman, G. (1999). XIV*—Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error. *Proceedings of the Aristotelian Society, 99*(1), 315-332. https://doi.org/10.1111/1467-9264.00062

Harman, G. (2000). The nonexistence of character traits. *Proceedings of the Aristotelian Society, 100*, 223-226. https://doi.org/10.1111/1467-9264.00077

Harman, G. (2009). Skepticism about character traits. *The Journal of Ethics, 13*(2/3), 235-242. https://doi.org/10.1007/s10892-009-9050-6

Heal, J. (1978). Common knowledge. *Philosophical Quarterly, 28*(111): 116-131. https://doi.org/10.2307/2219358

Hill Jr., T. E. (2012). *Virtue, rules, and justice: Kantian aspirations*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199692002.001.0001

Hindriks, F. (2009). Corporate responsibility and judgment aggregation. *Economics and Philosophy, 25*(2), 161-177. https://doi.org/10.1017/S0266267109990034

Hirshleifer, D., Low, A., & Teoh, S. H. (2012). Are overconfident CEOs better innovators? *The Journal of Finance, 67*(4), 1457-1498. https://doi.org/10.1111/j.1540-6261.2012.01753.x

Holroyd, J. (2012). Responsibility for implicit bias. *Journal of Social Philosophy, 43*(3), 274-306. https://doi.org/10.1111/j.1467-9833.2012.01565.x

House Committee on Transportation & Infrastructure. (2020). *The Boeing 737 MAX aircraft: Costs, consequences, and lessons from its design, development, and certification - preliminary investigative findings.* https://www.govinfo.gov/app/details/GOVPUB-Y4_T68_2-fb0f3812fefe3515ebcf3f4170fce64b

160

Huebner, B. (2011). Genuinely collective emotions. *European Journal for Philosophy of Science, 1*(1), 89-118. https://doi.org/10.1007/s13194-010-0006-2

Hume, D. (1779). An enquiry concerning human understanding. In *Essays and treatises on several subjects, Vol 2: Containing An enquiry concerning human understanding, A dissertation on the passions, An enquiry concerning the principles of morals, and The natural history of religion.* (pp. 3-212): Unknown Publisher.

Hurka, T. (2001). *Virtue, vice, and value*. Oxford University Press. https://doi.org/10.1093/0195137167.001.0001

Hursthouse, R. (1999). *On virtue ethics*. Oxford University Press. https://doi.org/10.1093/0199247994.001.0001

Hursthouse, R. (2007). Practical wisdom: A mundane account. *Proceedings of the Aristotelian Society, 106*, 285-309. https://doi.org/10.1111/j.1467-9264.2006.00149.x

Hursthouse, R. (2016). Virtue ethics. *Stanford encyclopedia of philosophy*. https://plato.stanford.edu/entries/ethics-virtue/

Hussain, I., Shu, R., Tangirala, S., & Ekkirala, S. (2018). The voice bystander effect: How information redundancy inhibits employee voice. *Academy of Management Journal, 62*(3), 828-849. https://doi.org/10.5465/amj.2017.0245

Isen, A. M., & Levin, P. F. (1972). Effect of feeling good on helping: Cookies and kindness. *Journal of Personality and Social Psychology, 21*(3), 384-388. https://doi.org/10.1037/h0032317

Jankélévitch, V. (1949). *Traité des vertus*. Bordas.

Johnson King, Z. A. (2020). Praiseworthy motivations. *Noûs, 54*(2), 408-430. https://doi.org/10.1111/nous.12276

Johnston, M. (1992). How to speak of the colors. *Philosophical Studies, 68*(3), 221-263. https://doi.org/10.1007/BF00694847

Jones, M. B. (2013). The home advantage in individual sports: An augmented review. *Psychology of Sport and Exercise, 14*(3), 397-404. https://doi.org/10.1016/j.psychsport.2013.01.002

Jones, T. (2007). Numerous ways to be an open-minded organization: A reply to Lahroodi. *Social Epistemology, 21*(4), 439-448. https://doi.org/10.1080/02691720701746565

Jost, L., & Wuerth, J. (2011). *Perfecting virtue: New essays on Kantian ethics and virtue ethics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511973789

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Bibliography

Kamtekar, R. (2004). Situationism and virtue ethics on the content of our character. *Ethics, 114*(3), 458-491. https://doi.org/10.1086/381696

Kitroeff, N., & Gelles, D. (2019, April 20). Claims of shoddy production draw scrutiny to a second Boeing jet. *The New York Times*. https://www.nytimes.com/2019/04/20/business/boeing-dreamliner-production-problems.html

Knobe, J., & Prinz, J. (2008). Intuitions about consciousness: Experimental studies. *Phenomenology and the Cognitive Sciences, 7*(1), 67-83. https://doi.org/10.1007/s11097-007-9066-y

Komite Nasional Keselamatan Transportasi. (2019). *Final aircraft accident investigation report* (Report No. KNKT.18.10.35.04). http://knkt.dephub.go.id/knkt/ntsc_aviation/baru/2018%20-%20035%20-%20PK-LQP%20Final%20Report.pdf

Konzelmann Ziv, A. (2012). Institutional virtue: How consensus matters. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 161*(1), 87-96. https://doi.org/10.1007/s11098-012-9933-4

Konzelmann Ziv, A., & Schmid, H. B. (2014). *Institutions, emotions, and group agents: Contributions to social ontology*. https://doi.org/10.1007/978-94-007-6934-2

Kopec, M. & Miller, S. (2018). Shared intention is not joint commitment. *Journal of Ethics and Social Philosophy 13*(2): 179-189. https://doi.org/10.26556/jesp.v13i2.250

Korsgaard, C. (1996). *The sources of normativity*. Cambridge University Press. https://doi.org/10.1017/CBO9780511554476

Kupperman, J. J. (2001). The indispensability of character. *Philosophy, 76*(296), 239-250. https://doi.org/10.1017/S0031819101000250

Kutz, C. (2000). Acting together. *Philosophy and Phenomenological Research 61*(1), 1-31. https://doi.org/10.2307/2653401

Lackey, J. (2016). What is justified group belief? *The Philosophical Review, 125*(3), 341-396. https://doi.org/10.1215/00318108-3516946

Lahroodi, R. (2007). Collective epistemic virtues. *Social Epistemology, 21*(3), 281-297. https://doi.org/10.1080/02691720701674122

Laker, B., Cobb, D., & Trehan, R. (2021). *Too proud to lead: How hubris can destroy effective leadership and what to do about it*. Bloomsbury Business.

Lederman, H. (2018a). Common knowledge. In M. Jankovic and K. Ludwig (Eds.) *Handbook of collective intentionality* (pp. 181-195). Routledge. https://doi.org/10.4324/9781315768571-18

Lederman, H. (2018b). Uncommon Knowledge. *Mind*, *127*(508), 1069-1105. https://doi.org/10.1093/mind/fzw072

Lederman, H. (2018c). Two paradoxes of common knowledge: Coordinated attack and electronic mail. *Noûs, 52*(4), 921-945. https://doi.org/10.1111/nous.12186

Levitt, B., & March, J. G. (1988). Organizational learning. *Annual Review of Sociology, 14*(1), 319-338. https://doi.org/10.1146/annurev.so.14.080188.001535

Lewis, D. (1969). *Convention: A philosophical study*. Wiley-Blackwell.

Liljenquist, K., Zhong, C.-B., & Galinsky, A. D. (2010). The smell of virtue: Clean scents promote reciprocity and charity. *Psychological Science, 21*(3), 381-383. https://doi.org/10.1177/0956797610361426

List, C. (2018). What is it like to be a group agent? *Noûs, 52*(2), 295-319. https://doi.org/10.1111/nous.12162

List, C., & Pettit, P. (2002). Aggregating sets of judgments: An impossibility result. *Economics and Philosophy, 18*(1), 89-110. https://doi.org/10.1017/S0266267102001098

List, C., & Pettit, P. (2004). Aggregating sets of judgments: Two impossibility results compared. *Synthese, 140*(1), 207-235. https://doi.org/10.1023/B:SYNT.0000029950.50517.59

List, C., & Pettit, P. (2011). *Group agency: The possibility, design, and status of corporate agents*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199591565.001.0001

Ludwig, K. (2007). Collective intentional behavior from the standpoint of semantics. *Noûs 41*(3), 355-393. https://doi.org/10.1111/j.1468-0068.2007.00652.x

Ludwig, K. (2014). Proxy agency in collective action. *Noûs, 48*(1), 75-105. https://doi.org/10.1111/nous.12013

MacIntyre, A. (1981). *After virtue: A study in moral theory*. University of Notre Dame Press.

Macpherson, W. (1999). *The Stephen Lawrence inquiry: Report of an inquiry by William Macpherson of Cluny*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/277111/4262.pdf

Bibliography

McDowell, J. (1979). Virtue and reason. *The Monist, 62*(3), 331-350. https://doi.org/10.5840/monist197962319

McFall, L. (1987). Integrity. *Ethics 98*, 5-20. https://doi.org/10.1086/292912

Meilaender, G. C. (1984). *Theory and practice of virtue*. Notre Dame Press.

Merritt, M., Doris, J. M., & Harman, G. (2010). Character. In J. M. Doris (Ed.), *The moral psychology handbook* (pp. 355-401). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199582143.003.0012

Milgram, S. (1974). *Obedience to authority: An experimental view*. Harper Collins.

Miller, C. (2003). Social psychology and virtue ethics. *The Journal of Ethics, 7*(4), 365-392. https://doi.org/10.1023/A:1026136703565

Miller, C. (2020). Motivation and the virtue of honesty: Some conceptual requirements and empirical results. *Ethical Theory and Moral Practice, 23*(2), 355-371. https://doi.org/10.1007/s10677-019-10055-1

Miller, S. (2001). *Social action: A teleological account*. Cambridge University Press. https://doi.org/10.1017/CBO9780511612954

Mischel, W. (1968). *Personality and assessment*. Wiley.

Mischel, W. (2009). From *Personality and assessment* (1968) to personality science. *Journal of Research in Personality, 43*(2), 282-290. https://doi.org/10.1016/j.jrp.2008.12.037

Moore, G. (2005). Corporate character: Modern virtue ethics and the virtuous corporation. *Business Ethics Quarterly, 15*(4), 659-685. https://doi.org/10.5840/beq200515446

Moore, G. (2015). Corporate character, corporate virtues. *Business Ethics: A European Review, 24*(2), 99-114. https://doi.org/10.1111/beer.12100

Mudde, A., Hoefnagels, C., Wijnen, L., & Kremers, S. (2007). Bystanding an abused child: Testing the applicability of the stages of change construct. *Health Education Research, 22*, 128-138. https://doi.org/10.1093/her/cyl056

Myers, D. G. (2008). *Social psychology*. McGraw-Hill.

Nagel, T. (1993). Moral luck. In D. Statman (Ed.), *Moral luck* (pp. 141-166). State University of New York Press.

Nelson, R. R., & Winter, S. G. (1985). *An evolutionary theory of economic change*. Belknap Press.

Nevill, A. M., & Holder, R. L. (1999). Home advantage in sport: An overview of studies on the advantage of playing at home. *Sports Medicine, 28*(4), 221-236. https://doi.org/10.2165/00007256-199928040-00001

Nevill, A. M., Holder, R. L., Bardsley, A., Calvert, H., & Jones, S. (1997). Identifying home advantage in international tennis and golf tournaments. *Journal of Sports Sciences, 15*(4), 437-443. https://doi.org/10.1080/026404197367227

Nozick, R. (1974). *Anarchy, State, Utopia*. Basic Books.

Nussbaum, M. (1990). *Love's Knowledge*. Oxford University Press.

Nussbaum, M. (2006). *Frontiers of justice*. Belknap Press. https://doi.org/10.2307/j.ctv1c7zftw

Odell, M. (2011, July 21). Boeing bows to pressure on 737 upgrade. *The Financial Times*. https://www.ft.com/content/cd0c2dac-b2d0-11e0-bc28-00144feabdc0

O'Hara, N. (2018). Implications for current metaethics. In N. O'Hara (Ed.), *Moral certainty and the foundations of morality* (pp. 155-190). Springer International. https://doi.org/10.1007/978-3-319-75444-4_6

Olin, L., & Doris, J. M. (2014). Vicious minds. *Philosophical Studies, 168*(3), 665-692. https://doi.org/10.1007/s11098-013-0153-3

Oreskes, N., & Conway, E. M. (2010). *Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming*. Bloomsbury Press.

Osnos, E. (2011). China's bystander effect. *The New Yorker*. https://www.newyorker.com/news/evan-osnos/chinas-bystander-effect

Ossei-Owusu, S. (2021). Police quotas. *New York University Law review, 96*, 529-604.

Oxfam International. (n.d.). Immediate response actions: Sexual misconduct. https://www.oxfam.org/en/immediate-response-actions-sexual-misconduct

Paternotte, C. (2011). Being realistic about common knowledge: A Lewisian approach. *Synthese, 183*(2), 249-276. https://doi.org/10.1007/s11229-010-9770-y

Paternotte, C. (2014). Constraints on joint action. In M. Gallotti, & J. Michael (Eds.), *Perspectives on social ontology and social cognition* (pp. 103-123). Springer. https://doi.org/10.1007/978-94-017-9147-2_8

Paternotte, C. (2015). The epistemic core of weak joint action. *Philosophy and Psychological Research, 28*(1), 70-93. https://doi.org/10.1080/09515089.2013.819614

Pentland, B. T., Hærem, T., & Hillison, D. (2011). The (n)ever-changing world: Stability and change in organizational routines. *Organization Science, 22*(6), 1369-1383. https://doi.org/10.1287/orsc.1110.0624

Pettit, P. (2004). Groups with minds of their own. In F. Schmitt (Ed.), *Socializing metaphysics*. Rowman and Littlefield.

Bibliography

Pettit, P. (2007). Responsibility incorporated. *Ethics, 117*(2), 171-201.
https://doi.org/10.1086/510695

Pieper, J. (1934). *Vom sinn der Tapferkeit*. Jakob Hegner-Verlag.

Pollard, R. (2006). Worldwide regional variations in home advantage in association football.
Journal of Sports Sciences, 24(3), 231-240.
https://doi.org/10.1080/02640410500141836

Pollard, R. (2008). Home advantage in football: A current review of an unsolved puzzle. *The
Open Sports Sciences Journal, 1,* 12-14.
https://doi.org/10.2174/1875399X00801010012

Price, A. W. (2011). *Virtue and reason in Plato and Aristotle*. Oxford University Press.
https://doi.org/10.1093/acprof:oso/9780199609611.001.0001

Price, H. H. (1954). Belief and will. *Aristotelian Society Supplementary Volume, 28*(1), 1-26.
https://doi.org/10.1093/aristoteliansupp/28.1.1

Prinz, J. (2009). The normativity challenge: Cultural psychology provides the real threat to
virtue ethics. *The Journal of Ethics, 13*(2/3), 117-144. https://doi.org/10.1007/s10892-
009-9053-3

Quinton, A. (1976). The presidential address: Social objects. *Proceedings of the Aristotelian
Society, 76*(1), 1-28. https://doi.org/10.1093/aristotelian/76.1.1

Railton, P. (1984). Alienation, consequentialism, and the demands of morality. *Philosophy
and Public Affairs, 13*(2), 134-171.

Rhodes, C. (2016). Democratic business ethics: Volkswagen's emissions scandal and the
disruption of corporate sovereignty. *Organization Studies, 37*(10), 1501-1518.
https://doi.org/10.1177/0170840616641984

Ritchie, K. (2020). Minimal cooperation and group roles. In A. Fiebich (Ed.), *Minimal
cooperation and shared agency* (pp. 93-109). Springer International.
https://doi.org/10.1007/978-3-030-29783-1_6

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the
attribution process. In L. Berkowitz (Ed.), *Advances in Experimental Social
Psychology* (Vol. 10, pp. 173-220). Academic Press. https://doi.org/10.1016/S0065-
2601(08)60357-3

Ross, L., & Nisbett, R. E. (2011). *The Person and the situation: Perspectives of social
psychology*. Pinter & Martin.

Russell, D. (2009). *Practical intelligence and the virtues*. Clarendon Press.
https://doi.org/10.1093/acprof:oso/9780199565795.001.0001

Sabini, J., & Silver, M. (2005). Lack of character? Situationism critiqued. *Ethics, 115*(3), 535-562. https://doi.org/10.1086/428459

Salomone-Sehr, J. (2022). Cooperation: With or without shared intentions. *Ethics*, *132*(2), 414-444. https://doi.org/10.1086/716877

Sanderson, C. A. (2020). *Why we act: Turning bystanders into moral rebels*. Belknap Press. https://doi.org/10.4159/9780674241817

Sandin, P. (2007). Collective military virtues. *Journal of Military Ethics, 6*(4), 303-314. https://doi.org/10.1080/15027570701755505

Schwitzgebel, E. (2015). If materialism is true, the United States is probably conscious. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 172*(7), 1697-1721. https://doi.org/10.1007/s11098-014-0387-8

Searle, J. (1990). Collective intentions and actions. In P. R. C. Jerry Morgan & M. Pollack (Eds.), *Intentions in communication* (pp. 401-415). MIT Press.

Shepardson, D. (2020). New Boeing chief executive takes over with 737 MAX crisis unresolved. https://www.reuters.com/article/boeing-737-max-ceo-calhoun-idUSL1N29I00J

Shoda, Y., Mischel, W., & Wright, J. C. (1994). Intraindividual stability in the organization and patterning of behavior: incorporating psychological situations into the idiographic analysis of personality. *Journal of Personality and Social Psychology, 67*(4), 674-687. https://doi.org/10.1037//0022-3514.67.4.674

Sidgwick, H. (1907). *The methods of ethics* (7th edition). Macmillan.

Sie, M., & van Voorst Vader-Bours, N. (2016). Stereotypes and prejudices: Whose responsibility? Personal responsibility vis-à-vis implicit bias. In M. Brownstein & J. Saul (Eds.), *Implicit bias and philosophy vol. 2* (pp. 90-114). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198766179.003.0005

Sims, R. R., & Brinkmann, J. (2003). Enron ethics (Or: Culture matters more than codes). *Journal of Business Ethics, 45*(3), 243-256. https://doi.org/10.1023/A:1024194519384

Slote, M. (1983). *Goods and virtues*. Oxford University Press.

Slote, M. (1995). Agent-based virtue ethics. *Midwest Studies In Philosophy, 20*(1), 83-101. https://doi.org/10.1111/j.1475-4975.1995.tb00306.x

Slote, M. (2001). *Morals from motives*. Oxford University Press. https://doi.org/10.1093/0195138376.001.0001

Bibliography

Smith, E. R., Seger, C. R., & Mackie, D. M. (2007). Can emotions be truly group level? Evidence regarding four conceptual criteria. *Journal of Personality and Social Psychology, 93*(3), 431-446. https://doi.org/10.1037/0022-3514.93.3.431

Snow, N. (2010). *Virtue as social intelligence: An empirically grounded theory*. Routledge Press. https://doi.org/10.4324/9780203880579

Solomon, R. C. (2003). Victims of circumstances? A defense of virtue ethics in business. *Business Ethics Quarterly, 13*(1), 43-62. https://doi.org/10.5840/beq20031314

Sosa, E. (1980). The raft and the pyramid: Coherence versus foundations in the theory of knowledge. *Midwest Studies In Philosophy, 5*(1), 3-26. doi:10.1111/j.1475-4975.1980.tb00394.x

Sosa, E. (2007). *A Virtue Epistemology*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199297023.001.0001

Sreenivasan, G. (2002). Errors about errors: Virtue theory and trait attribution. *Mind, 111*(441), 47-68. https://doi.org/10.1093/mind/111.441.47

Sreenivasan, G. (2020). *Emotion and virtue*. Princeton University Press. https://doi.org/10.23943/princeton/9780691134550.001.0001

Stalnaker, R. C. (2009). On Hawthorne and Magidor on assertion, context, and epistemic accessibility. *Mind* 118(470): 399-409. https://doi.org/10.1093/mind/fzp061

Staub, E. (2015). *The roots of goodness and resistance to evil: Inclusive caring, moral courage, altruism born of suffering, active bystandership, and heroism*. Oxford University Press.

Stoutland, S. (1997). Why are philosophers of action so anti-social? In L. Alanen, S. Heinämaa & T. Wallgren (Eds.), *Commonality and particularity in ethics*. Palgrave Macmillan. https://doi.org/10.1007/978-1-349-25602-0_3

Suarez, F. F., & Montes, J. S. (2019). An integrative perspective of organizational responses: Routines, heuristics, and improvisations in a Mount Everest expedition. *Organization Science, 30*(3), 573-599. https://doi.org/10.1287/orsc.2018.1271

Sunstein, C. (2014). *Why nudge? The politics of libertarian paternalism*. Yale University Press.

Swanton, C. (2003). *Virtue ethics: A pluralistic view*. Oxford University Press. https://doi.org/10.1093/0199253889.001.0001

Swanton, C. (2021). *Target centred virtue ethics*. Oxford University Press. https://doi.org/10.1093/oso/9780198861676.001.0001

Tanesini, A. (2018). Epistemic vice and motivation. *Metaphilosophy, 49*(3), 350-367. https://doi.org/10.1111/meta.12301

Thomas, G. (2006). What comes next? *Air Transport World, 43*(5), 51-52.

Thomson, J. J. (1997). The right and the good. *The Journal of Philosophy, 94*(6), 273-298. https://doi.org/10.2307/2564542

Tollefsen, D. P. (2015). *Groups as agents*. Polity Press.

Toner, C. (2014). The full unity of the virtues. *The Journal of Ethics, 18*(3), 207-227. https://doi.org/10.1007/s10892-014-9165-2

Tsoukas, H., & Chia, R. (2002). On organizational becoming: Rethinking organizational change. *Organization Science, 13*(5), 567-582. https://doi.org/10.1287/orsc.13.5.567.7810

Tuomela, R. (2000). *Cooperation: A philosophical study*. Springer. https://doi.org/10.1007/978-94-015-9594-0

Tuomela, R. (2007). *The philosophy of sociality: The shared point of view*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195313390.001.0001

Tuomela, R. (2013). *Social Ontology: Collective intentionality and group agents*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199978267.001.0001

Upton, C. L. (2005). A contextual account of character traits. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 122*(2), 133-151. https://doi.org/10.1007/s11098-004-6218-6

Upton, C. L. (2009a). The structure of character. *The Journal of Ethics, 13*(2/3), 175-193. https://doi.org/10.1007/s10892-009-9047-1

Upton, C. L. (2009b). Virtue ethics and moral psychology: The situationism debate. *The Journal of Ethics, 13*(2/3), 103-115. https://doi.org/10.1007/s10892-009-9054-2

Vallor, S. *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190498511.001.0001

Vanderschraaf, P. (2013). Common Knowledge. *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/common-knowledge/

Vogus, T. J., & Sutcliffe, K. M. (2007). Organizational resilience. *Proceedings of the 2007 IEEE International Conference on Systems, Man, and Cybernetics*, 3418-3422. https://doi.org/10.1109/ICSMC.2007.4414160

von Scheve, C., & Ismer, S. (2013). Towards a theory of collective emotions. *Emotion Review, 5*(4), 406-413. https://doi.org/10.1177/1754073913484170

Bibliography

Vranas, P. B. M. (2005). The indeterminacy paradox: Character evaluations and human psychology. *Noûs, 39*(1), 1-42. https://doi.org/10.1111/j.0029-4624.2005.00492.x

Washington, N., & Kelly, D. (2016). Who's responsible for this? Moral responsibility, externalism, and knowledge about implicit bias. In M. Brownstein & J. Saul (Eds.), *Implicit bias and philosophy vol. 2*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198766179.003.0002

Watson, G. (1990). On the primacy of character. In O. Flanagan & A. Rorty (Eds.), *Identity, character, and morality*. MIT Press. https://doi.org/10.7551/mitpress/3645.003.0026

Weitzman, H. (2011, April 27). Boeing dismisses Airbus threat. *The Financial Times*. https://www.ft.com/content/765db61e-70c8-11e0-9b1d-00144feabdc0

West, R. (2018). Virtue ethics is empirically adequate: A defense of the CAPS response to situationism. *Pacific Philosophical Quarterly, 99*(1), 79-111. https://doi.org/10.1111/papq.12196

Wielenberg, E. J. (2006). Saving character. *Ethical Theory and Moral Practice, 9*(4), 461-491. https://doi.org/10.1007/s10677-006-9010-1

Wiggins, D. (1998). Natural and artificial virtues. In R. Crisp (Ed.), *How should one live? Essays on the virtues* (pp. 131-140). Oxford University Press. https://doi.org/10.1093/0198752342.003.0008

Williams, B. (1981). *Moral luck*. Cambridge University Press. https://doi.org/10.1017/CBO9781139165860

Williams, B. (1985). *Ethics and the limits of philosophy*. Harvard University Press. https://doi.org/10.4324/9780203969847

Williams, B. (2002). *Truth and truthfulness: An essay in genealogy*. Princeton University Press. https://doi.org/10.1515/9781400825141

Wood, W. J., & Roberts, R. C. (2007). *Intellectual virtues: An essay in regulative epistemology*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199283675.001.0001

Wright, J. C., & Mischel, W. (1987). A conditional approach to dispositional constructs: the local predictability of social behavior. *Journal of Personality and Social Psychology, 53*(6), 1159-1177. https://doi.org/10.1037//0022-3514.53.6.1159

Zagzebski, L. (1996). *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge University Press. https://doi.org/10.1017/CBO9781139174763

Zagzebski, L. (2010). Exemplarist virtue theory. *Metaphilosophy, 41*(1-2), 41-57. https://doi.org/10.1111/j.1467-9973.2009.01627.x

Zagzebski, L. (2013). Moral exemplars in theory and practice. *Theory and Research in Education, 11*(2), 193-206. https://www.doi.org/10.1177/1477878513485177

Zimbardo, P. (1973, April 8). The mind is a formidable jailer: A Pirandellian prison. *The New York Times Magazine*.

About the author

Barend Cornelis de Rooij was born in Tilburg, the Netherlands, on March 20, 1992. He obtained his BA (Hons) degree in Liberal Arts & Sciences in 2013 at Leiden University College (*summa cum laude*), graduating top of his class. During the last year of his bachelor studies, he spent one semester abroad studying philosophy at Rutgers University in New Jersey (4.0 GPA). He continued his education at the University of St. Andrews, where he obtained an MLitt degree in Philosophy (with distinction overall) in 2015. At the Catholic University of Leuven, De Rooij obtained an MA degree in European Studies (*magna cum laude*) in 2016 and an MA degree by research in Philosophy (*summa cum laude*) in 2017. He joined the NWO-funded project 'Towards Professional Epistemic Justice: Finance and Medicine' in the same year, exploring topics of justice, epistemology, and group agency under the supervision of Prof. Boudewijn de Bruin (Rijksuniversiteit Groningen) and Prof. Miranda Fricker (New York University).

During his doctoral studies, De Rooij spent the first two years at the University of Sheffield, where he supervised tutorials in epistemology, ethics, and logic. He spent the last the last two years at the Rijksuniversiteit Groningen, where he taught multiple courses in business ethics. As of January 2022, De Rooij works as a lecturer in philosophy at Tilburg University.

Abstract

We often speak of businesses, institutions, and other organizations as if they have character traits of various kinds. Say, for instance, that a pharmaceutical company has a history of quadrupling the price of essential medication without apparent justification. While we may criticize that organization for the individual *acts* of raising its prices, we may also be tempted to criticize it for being the kind of organization that puts profits before people; that is, we may blame it for having a greedy *character*. On the other hand, a pharmaceutical company that waves lucrative patents in order to make life-saving drugs more widely available may instead deserve praise for its generous character.

In the history of philosophy, such character-based evaluations have been closely associated with the tradition of virtue ethics, and, more recently, virtue epistemology. Both virtue ethicists and epistemologists study the features of our character that make us excellent (*virtues*) or deficient (*vices*) in various respects. But while theorists working within the fields of virtue ethics and epistemology typically conceive of virtuous and vicious character traits as features belonging to *individuals*, they have given far less thought to how concepts of vice and virtue may apply to *groups*.

This dissertation aims to bridge the gap between virtue theory and the ethics and epistemology of groups by investigating the notion of *group character*. As such, it takes existing research on the ethics of groups in new directions by critically investigating the conceptual and empirical foundations of an agent-centered approach to group moral theory and epistemology. The overarching research question this dissertation seeks to answer is whether, and how, group agents are capable of manifesting virtuous or vicious character traits as subjects in their own right. Answering the *whether*-question with a resounding *yes*, it aims to develop an account of group character that is both conceptually rigorous and empirically sound.

On the account of group character, virtue, and vice developed here, group agents instantiate a collectively virtuous or vicious character trait if they are organized so as to *function* in a virtuous or vicious way. This requires realizing a virtuous or vicious functional state at the collective level. Thus, say, a pharmaceutical company qualifies as collectively

greedy if it is organized in such a way that it is collectively disposed to pursue profits at all costs. Given that different virtues and vices correspond to different functional states, and given that these functional states can be realized by way of many different organizational structures, this dissertation shows that there are innumerable ways in which group agents can realize collective virtues and vices.

Over the course of five chapters, this dissertation makes this functionalist account of group character more concrete, illustrates its central applications, and defends it from influential objections. *Chapter one* sets the stage for answering the main research question by critically examining the current state of virtue theory. Though virtue ethicists have gone farthest in their exploration of virtuous or vicious character traits, it shows that character-based evaluations have also made an appearance in other moral traditions. Across these traditions, the power of the concepts of virtue and vice lies in their ability to represent the worth of an agent's character in a particular way. This chapter focuses on two conceptions of virtue and vice in particular. On *ends-based* conceptions of virtues and vice, virtuous or vicious character traits are traits that dispose their possessor to realize virtuous or vicious ends (or consequences). On *motivation-based* conceptions, virtues and vices additionally require the presence of having virtuous or vicious motivation.

Chapter two aims to show how group agents of various kinds may qualify as virtuous or vicious on both conceptions of virtue and vice. As a starting point, it considers existing accounts of collective virtue and vice based in Margaret Gilbert's plural subject account of group agency. Though these accounts offer compelling analyses of *some* collective virtues and vices, this chapter argues that they suffer from two limitations that restrict their applicability to *motivation*-based virtue and vice. In order to accommodate group virtues and vices of all kinds, it subsequently develops an entirely novel, *functionalist* account of collective virtue and vice. As we have seen, this view entails that collective virtues and vices arise when corporations or other group actors are organized so to as to *function* as virtuous or vicious agents. This new account of collective virtue and vice can model a broad range of collective character traits, has considerable empirical purchase, and ties in well with empirical literature from the organizational sciences.

Chapter three deals with an influential objection against the virtue theoretical project, which has come to be known as *the situationist challenge*. Situationists charge that the appeal of virtue theory rests on problematic commitments that do not hold up to empirical scrutiny. Specifically, they argue that empirical research shows that human behavior is highly vulnerable to (often trivial) situational influences, failing to display the cross-situational

consistency required for genuine virtues and vices. This chapter critically examines the implications of the situationist challenge for a functionalist account of group virtue, expanding the situationist challenge in two ways. First, it considers what situationist research tells us about the impact of group dynamics on individual virtues and vices. Second, it investigates whether group *agents* are vulnerable to situationist critiques as well. While one may be tempted to conclude that group agents are exempt from the situationist challenge because they do not face the same psychological limitations as individual agents, this chapter nevertheless shows that groups are liable to situational influences much like individuals are. Fortunately, this does not invalidate a functionalist account of group virtue. This is because we can defuse the situationist challenge by highlighting the importance of *local* behavioral regularities.

Chapter four illustrates the practical relevance of a functionalist account of collective virtue and vice, focusing on virtues and vices of an *epistemic* nature. It does this by way of a case study of the American aircraft manufacturer Boeing. Drawing on media reports and official investigations, it analyzes the *corporate epistemic culture* that shaped the conditions leading up to the deadly crashes of two of its Boeing 737 Max jets in 2018 and 2019. Though it would be a mistake to say that these crashes were entirely to blame on epistemic problems, the chapter concludes that collective epistemic vice undoubtedly played a part. The chapter also includes recommendations for fostering an environment conducive to collective epistemic virtue instead.

Finally, in *chapter five*, this dissertation investigates the role of *improvisation* in accounts of collective virtue and vice. While groups are capable of realizing many of the conditions required for the possession of virtue or vice, the existing literature on group agency does not capture all of these prerequisites. On most conceptions of virtue, for instance, cultivating a virtuous character requires that one is capable of improvising morally good behavior. As this chapter argues, however, extant accounts of joint action struggle to explain the possibility of jointly improvised action. The problem lies in the widely held view that jointly intentional behavior requires common knowledge of shared intentionality. In paradigmatic examples of improvisation, groups are able to improvise joint acts without prior knowledge of shared intentionality. The chapter proposes that we can resolve this problem by dropping the condition that common knowledge is required for joint action.

Nederlandse samenvatting

Het karakter van groepen – Een studie van collectieve (on)deugden

We spreken in ons alledaags taalgebruik regelmatig over het *karakter* van bedrijven, instituten, en andere organisaties. Beeldt u zich bijvoorbeeld in dat een farmaceutisch bedrijf de prijzen van een levensreddend medicijn zonder gegronde reden drastisch verhoogt. In eerste instantie kunnen we dat bedrijf bekritiseren voor de prijsverhoging op zich (d.w.z., *qua* handeling). Als echter blijkt dat het bedrijf in kwestie de prijzen van haar producten regelmatig zonder legitieme reden verhoogt, kunnen we ons ook uitspreken over het karakter van het bedrijf *an sich*. Kennelijk hebben we hier te maken met een *hebzuchtig* bedrijf – een bedrijf dat winsten hoger in het vaandel heeft staan dan het welzijn van haar consumenten. Het komt natuurlijk ook voor dat we een bedrijf of ander soort organisatie prijzen voor positieve karaktereigenschappen. Mocht een farmaceutisch bedrijf er bijvoorbeeld naar streven om levensreddende medicijnen tegen de kostprijs beschikbaar te stellen voor ontwikkelingslanden, kunnen we dat bedrijf prijzen voor haar liefdadigheid.

Binnen de filosofie worden dit soort beoordelingen op basis van karaktereigenschappen voornamelijk bestudeerd in de deugdethiek. Recentelijk kan daar de deugdepistemologie aan worden toegevoegd. Zowel deugdethici als deugdepistemologen doen onderzoek naar de karaktereigenschappen die ons *deugdzaam* dan wel *ondeugdzaam* maken. Ethici richten zich daarbij vooral op karaktereigenschappen die ons *moreel* prijzenswaardig maken (zoals liefdadigheid), terwijl epistemologen focussen op de karaktertrekken van een goed kenner of denker (zoals ruimdenkendheid). Hoewel de deugdtheorie op dit moment erg in trek is, bestudeert zij hoofdzakelijk de karaktereigenschappen van *personen*, niet die van de bedrijven, instituten, en andere organisaties wiens karaktereigenschappen we ook regelmatig prijzen of bekritiseren.

Dit proefschrift streeft ernaar om dit gat in de deugdtheorie te dichten door het karakter van bedrijven, instituten, en andere organisaties grondig onder de loep te leggen. De hoofdvraag die dit proefschrift zich daarbij stelt is of groepen in staat zijn om (on)deugdzame

karaktereigenschappen te realiseren, en, zo ja, op welke wijze. Een belangrijke doelstelling van het huidige onderzoek is om de groeps- en bedrijfsethiek te verbreden met een benadering waarin het *karakter* van groepen centraal staat. Om deze benadering te realiseren ontwikkelt dit proefschrift een nieuwe theorie over groepskarakter die zowel conceptueel als empirisch gezien sterk in elkaar zit.

Het concept van groepskarakter dat in dit proefschrift wordt ontwikkeld is geïnspireerd door het *functionalisme* binnen de hedendaagse filosofie van de geest. Concreet houdt dit concept in dat een groep een collectieve deugd of ondeugd tot stand brengt wanneer zij op collectief niveau op een deugdzame of ondeugdzame wijze *fungeert*. Specifiek is het hierbij van belang dat de organisatie van een groep gekenmerkt wordt door een (on)deugdzame *functionele toestand*. Een farmaceutisch bedrijf voldoet hier bijvoorbeeld aan wanneer zij omwille van haar bedrijfsstructuur de dispositie heeft om op groepsniveau hebzuchtige besluiten te nemen. Gezien het feit dat iedere (on)deugdzame karaktereigenschap bestaat uit een andere functionele toestand, en gezien het feit dat deze functionele toestanden vaak door meerdere organisatiestructuren gerealiseerd kunnen worden, kunnen collectieve (on)deugden echter ook op vele andere manieren tot stand worden gebracht.

Dit proefschrift bestaat uit vijf hoofdstukken waarin deze functionalistische theorie over het karakter van groepen tot in detail wordt uitgewerkt. Daarbij wordt ook stilgestaan bij een invloedrijk bezwaar tegen de deugdtheorie en de praktische relevantie van een deugdtheorie toegepast op bedrijven, instituten, en andere organisaties. Het *eerste hoofdstuk* bespreekt de huidige staat van de deugdethiek aan de hand van een kritische analyse van recent onderzoek naar (on)deugdzame karaktereigenschappen. Uit deze analyse komt naar voren dat we deze eigenschappen op grofweg twee manieren kunnen duiden. Allereerst kunnen we (on)deugden duiden aan de hand van de doelen die deze eigenschappen dienen of de consequenties die ze voortbrengen. Daarnaast kunnen we deze eigenschappen duiden op basis van de *motivatie* die ze behelzen.

Hoofdstuk twee begint de zoektocht naar een goed onderbouwde theorie over het karakter van groepen. Het doel van dit hoofdstuk is om aan te tonen dat ook groepen beide vormen van (on)deugdzame karaktereigenschappen kunnen realiseren. Deze zoektocht begint bij de huidige literatuur over collectieve deugden en ondeugden, met name de literatuur die zich baseert op het werk van Margaret Gilbert. Uit dit hoofdstuk zal blijken dat deze literatuur er niet in slaagt om uit te leggen hoe groepen een deugdzame *motivatie* kunnen hebben. Ook toont dit hoofdstuk aan dat een *functionalistische* theorie over collectieve (on)deugden dat wèl kan. Zoals we al zagen, stelt de nieuwe theorie die dit proefschrift ontwikkelt dat een groep

een collectieve deugd of ondeugd realiseert wanneer zij op collectief niveau een deugdzame of ondeugdzame functionele toestand teweegbrengt. Hoofdstuk twee bespreekt ook andere voordelen van deze theorie; namelijk het feit dat deze theorie empirisch gerechtvaardigd lijkt, een breed scala aan collectieve karaktereigenschappen kan modelleren, en goed aansluit op de literatuur binnen de organisatiewetenschappen.

Hoofdstuk drie bespreekt een invloedrijk bezwaar tegen de deugdtheorie en onderzoekt of dit bezwaar ook het concept van collectieve (on)deugd in de problemen brengt. De tegenwerping in kwestie komt uit de hoek van het *situationisme*, een belangrijke stroming binnen de psychologie. Volgens situationisten blijkt uit empirisch onderzoek dat ons gedrag onderhevig is aan (vaak triviale) situationele invloeden, en dus niet voortvloeit uit consistente karaktereigenschappen zoals deugden of ondeugden. Hoewel dit hoofdstuk beargumenteert dat de deugdtheorie zich wel degelijk tegen het situationisme wapenen, betoogt het ook dat de uitdaging waarvoor situationisten ons stellen breder is dan aanvankelijk werd gedacht. Zo hebben veel van de situationele invloeden die binnen het situationisme worden besproken een *collectieve* dimensie die onderbelicht is gebleven. Groepen *zelf* blijken bijvoorbeeld ook onderhevig te zijn aan situationele invloeden, hoewel de 'psychologie' (cognitieve infrastructuur) van groepen radicaal anders is dan die van individuen. Gelukkig betekent dit niet dat het concept van collectieve (on)deugd empirisch onhoudbaar is. Grofweg dezelfde strategieën die deugdethici hebben gebruikt om het concept van individuele (on)deugd te verdedigen zijn ook toepasbaar op de collectieve variant.

Hoofdstuk vier illustreert de praktische relevantie van een functionalistische analyse van het karakter van groepen aan de hand van een recente casus. Deze illustratie richt zich vooral op collectieve deugden en ondeugden met een epistemisch karakter; dat wil zeggen, op de eigenschappen die groepen effectief maken (of juist niet) als het aankomt op het opdoen, verwerken, en communiceren van kennis en informatie. De casus die onder de loep wordt genomen betreft de Amerikaanse vliegtuiggigant Boeing, wiens 737 Max vliegtuigen in 2018 en 2019 betrokken waren bij twee dodelijke ongelukken. Uit de analyse die dit hoofdstuk maakt blijkt dat er op epistemisch vlak veel aan te merken was op de bedrijfsvoering van Boeing en dat de vliegtuigbouwer zich op collectief niveau epistemisch gezien ondeugdzaam heeft opgesteld. Hoewel dit niet betekent dat de dodelijke ongelukken volledig te herleiden zijn naar een collectief ondeugdzaam karakter, lijkt dit karakter wel degelijk een rol te hebben gespeeld. Hoofdstuk vier bespreekt ook welke stappen een organisatie kan ondernemen op een epistemisch gezien *deugdzaam* karakter tot stand te brengen.

Hoofdstuk vijf sluit af door het belang van *groepsimprovisatie* te bespreken. De voornaamste deugdtheorieën stellen namelijk dat deugdzaam handelen veronderstelt dat een deugdzaam persoon moreel (of epistemisch) prijzenswaardig handelen kan improviseren. Het blijkt echter dat veel analyses van *groeps*handelingen niet kunnen verklaren hoe *groeps*actoren kunnen improviseren. Dat komt omdat deze analyses stellen dat groepshandelingen uitsluitend mogelijk zijn als groepsleden gemeenschappelijke kennis hebben van elkaars intenties, doelstellingen, of verwachtingen; iets waarvan geen sprake is bij handelingen die geïmproviseerd zijn en dus spontaan tot stand komen. Om dit probleem op te lossen beargumenteert dit hoofdstuk dat groepen ook in staat kunnen zijn om samen te handelen zonder gemeenschappelijke kennis van elkaars intenties, doelstellingen, of verwachtingen. In plaats daarvan spelen andere epistemische factoren een rol, zoals het praktisch inzicht waardoor groepsleden in staat zijn hun handelen gezamenlijk te coördineren.