LOYOLA UNIVERSITY CHICAGO

COLLECTIVE RESPONSIBILITY BY AGREEMENT

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE GRADUATE SCHOOL
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

PROGRAM IN PHILOSOPHY

BY

DAVID ATENASIO

CHICAGO, IL

MAY 2019
ACKNOWLEDGEMENTS

Upon hearing a summary of my project, Richard W. Miller described it as “lawyerly.” I have no idea if he meant it as a compliment, but whatever lawyerly virtues this dissertation exemplifies are in no small part due to the encouragement, advice and criticism of my excellent dissertation committee: Thomas Carson, Joy Gordon and Heidi Malm. I have also benefitted from remarks by and discussions with conference attendees of the Great Lakes Philosophy Conference, Illinois Philosophical Association Conference, Central States Philosophical Association Conference, American Philosophical Association: Central Division Meeting and Ethics and Political, Social and Legal Annual Conference, who graciously suffered through earlier versions of chapters from this dissertation. I must also thank the faculty members and graduate students of Loyola’s philosophy department for their support and conversations. I am fortunate enough to say that there are too many of you to list individually. Finally, I have benefitted greatly from reviewer comments made by anonymous referees at Ethical Theory and Moral Practice, Res Publica and Social Theory and Practice. I do not know your names, but your efforts are not forgotten.

The life of the mind requires material support, and on that measure I have been quite blessed. I have been able to write, teach and research in security because of the excellent funding and benefits package offered to me by Loyola University Chicago from the 2013-2014 to 2017-2018 academic years. For the current 2018-2019 year, I must thank the Andrew W. Mellon Foundation and American Council of Learned Sciences for providing such generous funding by
way of a Dissertation Completion Fellowship. It is an honor to have been chosen for such a
prestigious award.

Finally, I would like to thank my wife Naomi for her love, patience and support and my
son Michael for his persistent efforts at stealing my mouse and throwing my international law
books on the floor.
In those days they shall no longer say: ‘The fathers have eaten sour grapes, and the children’s teeth are set on edge.’

—Jeremiah 31:29
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

INTRODUCTION 1

CHAPTER ONE: COLLECTIVE RESPONSIBILITY AND TACIT AUTHORIZATION 12

Responsibility and Authorization 16
Tacit Authorization 32
The Moral Separateness of Agents and Proportionality 55
What Fairness Requires 60
Conclusion 64

CHAPTER TWO: COLLECTIVE ACTION, COLLECTIVE RESPONSIBILITY 67

From Individual to Collective Action 71
From Collective Intentions to Collective Responsibility 76
Participatory Intentions and Proportionate Blame 80
Intentions, Intuitions and the Complicity Principle 85
Conclusion 93

CHAPTER THREE: BLAMELESS PARTICIPATION IN STRUCTURAL INJUSTICE 96

Jaspers and Arendt on Political Responsibility 98
Participation in Structural Injustice 103
Disambiguating Structural Injustice 111
Systematically Unfair Structures 112
Structures That Burden an Unconsenting Third Party 116
Structures that Inadequately Distribute Resources 120
Objections 123
Conclusion 133

CHAPTER FOUR: NEGLIGENCE AND INTENTIONAL WRONGDOING 135

Negligence 137
May’s Social Existentialism 140
The Standard Liberal View 143
Four Objections to May’s Theory 147
Objections to the Standard Liberal View 153
Conclusion 161

CHAPTER FIVE: INDIVIDUAL LIABILITY FOR COLLECTIVE CRIMES 163

From the IMT to the ICC 165
Yamashita and Command Responsibility 169
Tadić and Joint Criminal Enterprise 173
The ICC and the Control Standard 178
Justice or Collective Punishment? 182
Conclusion 199

BIBLIOGRAPHY 202
INTRODUCTION

We live in a world of increasing social complexity. New methods of communication and organization allow us to achieve formerly unthinkable goals. This power can be a source of good: international aid organizations quickly mobilize to direct funds toward countries in crisis; social media provides an avenue for solidarity, friendship and coordination across geographical boundaries; fledgling institutions of global governance represent an important step in the process of restraining sovereign misconduct through the rule of law. But this heightened capacity for social organization has downsides. The First and Second World Wars demonstrated the possibility of war on a global scale. Political decisions and financial policies have led to manmade famines and staggering inequalities (Pogge, 2008, pp. 118-122; Sen, 1981, ch. 6). Many fear that the rise of modern industry will have devastating consequences for the integrity of our shared environment (R. Miller, 2010, ch. 4).

When surveying these outcomes, it can be difficult to articulate who we ought to blame or praise. In the West, we largely inherit a moral framework rooted in the tradition of medieval Christianity and early modern philosophy. This tradition is liberal and individualistic. Western ethics values primarily the autonomy and dignity of the individual agent. According to this view, an individual’s merit increases or decreases based on the actions that she intentionally carries out. When her individual conduct is good, she is credited, and when her conduct is bad, she is blamed; we cannot rightfully blame one individual for the faulty conduct of another without a
good reason. Western ethics also places an emphasis on proportionality. Not only must we primarily credit or discredit an individual for her own intentional conduct, we should only do so in proportion to the magnitude of her wrongdoing or success. An individual only deserves minimal blame for a minor moral infraction. However, if the transgression is serious enough, the individual deserves harsher blame and possibly legal forms of punishment.

The growing impact of collective action has led some to wonder if this liberal conception of morality is too facile. A strongly individualist conception of morality may have sufficed for a simpler age, but in the era of the corporation, we need an expanded conception of moral responsibility, a theory of collective moral responsibility (Cooper, 1968; French, 1984, viii-ix; Gomperz, 1939, p. 332; Isaacs, 2006; May, 1987, p. 2). A theory of collective responsibility aims to ascribe praise or blame to groups or group members for the collective actions that they intentionally carry out.

The idea of collective responsibility is hardly a radical evolution in moral thinking. In many older Western societies, officials used legal notions of collective responsibility to encourage self-policing where formal policing was not possible. Joel Feinberg notes that during the time of Christian feudalism in England, lawmakers developed a system of “compulsory universal suretyship” (1968, p. 679). Under universal suretyship, every individual was assigned to a neighborhood group. Members of a neighborhood group became insurers of each other’s conduct. When a member of a neighborhood group performed a wrongful deed, lawmakers would look to the neighborhood to produce the offending individual for punishment. If the neighborhood group did not produce the offending individual, lawmakers would fine the neighborhood group as a whole and demand that all group members compensate the victim (p. 680).
History offers up other examples of collective responsibility utilized as a tool of law or morals. Terrorists frequently justify their actions by appeal to the collective responsibility that their victims bear for group transgressions (Reiff, 2008, p. 228). Hitler and many of his followers believed that the whole Jewish race was responsible for the “catastrophic splintering’ and weakening of the German world” (Cooper, 2003, p. 17). After the conclusion of the U.S. Civil War, Chief of Staff Henry W. Halleck claimed that “it is a general law of war, that communities are accountable for the acts of their individual members” (as qtd in Darcy, 2007a, p. 9). There are more mundane examples as well. Teachers sometimes find it effective to threaten an entire class with punishment for the actions of a single student. For instance, a teacher might tell her class that if a single student talks during a lecture, the whole class will miss recess (Lipnowski, 1993, p. 122).

The demands of liberal principles of fairness and collective responsibility pull in opposite directions. If morality focuses too narrowly on the individual and her isolated actions, it risks missing the bigger picture of how the individual’s actions coordinate with and contribute to larger collective undertakings. If morality takes on too collective a character, patently innocent individuals will be held to blame for conduct that is completely outside of their locus of control or influence. Such a result would risk jettisoning much of what is valuable and rationally defensible in the traditional liberal moral project.

Two primary avenues of investigation emerge as important in the literature on collective responsibility. First, some theorists seek to discover which entities are capable of bearing moral responsibility. There are those who argue that moral responsibility only attaches to individual agents and never to collective entities. These theorists traditionally refer to themselves as “methodological individualists” (Copp, 1984, p. 251). Others contend that certain groups
themselves are capable of bearing moral responsibility. They would claim that it is not necessarily the executives at BP who are morally responsible for polluting the Gulf Coast. Rather, the corporation BP itself is to blame for its negligence. The notion that a single corporate entity may bear moral responsibility is typically called “corporate responsibility.”

The metaphysical issue of which entities might bear moral responsibility is a topic that has received much attention, but it is one that I set aside for the sake of this project. As much as possible, I will avoid philosophical concerns relating to the social ontology of groups. I will therefore stay agnostic as to whether it is correct to blame corporate entities as a whole. My interest lies in the second avenue of investigation that emerges in the literature on collective responsibility: the question of how one individual becomes morally responsible for the contributions another makes to a group action, for it is this question that strains at the limits of traditional ideals of fairness. Typically, we do not hold one individual responsible for the actions of another. But this is not always so. When a husband hires a hitman to kill his wife, we hold the husband responsible not only for conspiring to kill his wife, but also for the wife’s death, if the hitman is successful. We therefore fault the husband for the intentional actions of the contract killer. It is this distribution of responsibility to group members that I will term (distributive) collective responsibility (see Sverdlik, 1987, p. 62).

Theorists have offered up a number of solutions to explain how one individual might become responsible for the actions of her fellow group members. One popular method of distributing collective responsibility among contemporary analytic philosophers is by appeal to

1 H. Gomperz (1939, p. 332) refers to such responsibility as “social” responsibility.

2 American courts frequently assess harsher penalties on the instigators or authorizers of crime than on the proxy agents who do the dirty work; see Lussenhop (2015).
intentions or intentional states (Gadirov, 2011, p. 11; Gilbert, 2006, p. 99; Kutz, 2000a, p. 139; S. Miller, 2001, p. 65; Sadler, 2006; Sverdlik, 1986, p. 66). The intention model of collective responsibility establishes that one becomes responsible for the actions of one’s co-contributors by having a peculiar sort of intention to participate in wrongdoing. A collective intention is a mental state (or plurality of mental states) that aims at bringing about a collective aim with other agents. Some argue that the mental state is a *sui generis* form of “we-intention” (Gilbert, 2006, p. 100; Searle, 1990; Tuomela, 2006), while others insist that a collective intention is an individual intention whose content is a collective activity (see Kutz, 2000a, p. 108). In either case, an intention model of collective responsibility stipulates that an intention aimed at collective wrongdoing is at least necessary and often sufficient to make one responsible for the contributions of one’s fellow group members in some way.

Other theorists (May, 1992; Young, 2011) argue for an existentialist, “social connection” model of collective responsibility. A social connection model posits that we bear responsibility for more than our deliberate intentional actions. We also bear responsibility on account of the social relationships we have to others in our community and beyond. Young argues that we bear forward-looking responsibility for remediating and transforming the dispersed, structural harms in which we participate. We bear this responsibility whether or not our participation was informed and intentional (Young, 2011, p. 96). Larry May argues that we bear responsibility for all actions carried out by our fellow community members who share our hateful beliefs and attitudes (May, 1992, pp. 46-52). If an individual develops or fails to relinquish racist attitudes, then she may be held responsible for racist actions carried out by other members of her community who share her harmful attitudes (pp. 47-8).
In this dissertation, I propose an alternative to the current intention and social connection theories of distributive collective responsibility. I argue that we should understand the distribution of collective responsibility as a function of authorization. Usually, a person individually commits wrongdoing and we hold her to blame accordingly. However, sometimes a person authorizes others to commit wrongdoing on her behalf. When a person authorizes an agent to carry out wrongdoing, we are justified in holding the authorizer to blame for the wrongdoing her agent carries out on her behalf. Distributive collective responsibility is therefore a function of the authorization agreements we form with others to pursue our aims.

A theory of collective responsibility founded only on express authorization agreements will not extend very far. Most examples of collective wrongdoing typically rely on few if any express authorization agreements. I therefore extend an authorization account of distributing collective responsibility by introducing the notion of tacit authorization. I claim that any individual who performs certain intentional actions and meets certain criteria as part of a collective tacitly authorizes her fellow group members to act on her behalf. If a person tacitly authorizes her fellow group members to act in pursuit of a shared aim or goal, we can justifiably credit or discredit her for the relevant contributions her fellow group members contribute toward the shared aim or goal.

My primary aim in this dissertation is to develop a successful authorization model of distributing collective responsibility and argue that it is superior to prominent intention and social connection theories. In the first chapter, I articulate an authorization theory of collective responsibility. I argue that, given certain indications, it is reasonable to suppose that a group

---

3 Authorization theories of distributing collective responsibility are largely ignored or dismissed in the literature; see Atenasio (2018). However, Bazargan-Forward has begun to develop his own authorization account in (2017).
member tacitly authorizes her fellow group members to act on her behalf. These indications are: (1) a person must freely, successfully and intentionally make herself a member of the group or common plan,\(^4\) (2) the person must have accurate beliefs about the goals, purposes and general activities of the group or common plan and (3) the person must freely and intentionally make a substantial contribution to the relevant collective action. Once these criteria are met, it is reasonable to suppose that a group member tacitly authorizes her fellow group members to act on her behalf. I expand on the specifics of (1) through (3) and defend these criteria from some objections.

Also in chapter one, I provide concrete definitions for some terms that are used rather loosely in different moral and legal contexts. First, I define moral responsibility primarily as backward-looking fault. I differentiate moral responsibility as such from remedial or forward-looking responsibility. Second, I suggest clear definitions for authorization and consent. Frequently, “consent” is used as a catchall phrase for many different forms of social agreement. For the sake of clarity, I distinguish permissive consent from other forms of agreement. I define consent as the expression of willingness (in the proper context) to permit an intrusion into one’s personal sphere of autonomy. I define authorization as the expression of willingness (in the proper context) to have someone carry out an action on one’s behalf or as one’s proxy agent.

Once I articulate an authorization account of distributing collective moral responsibility, I proceed to argue that we ought to prefer it to other theories of distributing collective responsibility. One normative reason in favor of an authorization theory is that it can meet

\(^4\) The membership stipulation may also be met by individuals who were conscripted into a common plan, assuming that they embrace their roles and have the opportunity to leave the group without risking substantial harm to themselves.
rigorous demands of fairness. I propose two criteria of fairness, the separateness of agents and proportionality, and demonstrate that an authorization account satisfies both criteria. In the proceeding chapters, I argue that competing accounts either struggle to satisfy one or both of the criteria or are subject to conceptual objections.

In the second chapter, I argue for the superiority of an authorization account of distributing collective moral responsibility over prominent intention theories. I first canvas some developments in action theory and articulate how discussions of collective action have transitioned to discussions of collective responsibility. I then present the most prominent intention theory in the literature: Christopher Kutz’s (2000a) theory of participatory intentions. Kutz argues that moral responsibility for a collectively produced outcome distributes to all members who intentionally participated in bringing it about. I provide some reasons for why theorists have found Kutz’s account attractive. However, I argue that Kutz’s theory of participatory intentions leads to unfair and objectionable results. This presses us to look to other intention-state theories of distributing collective responsibility. However, I further argue that attempts to modify Kutz’s account of distributing collective responsibility also fail, as they provide no compelling rationale for theorists to accept the principle that sometimes individuals may be held to blame for the free and intentional contributions made by others toward a collective or shared aim.

In chapter three, I focus on Young’s (2011) social connection theory of responsibility. Young introduces a theory of responsibility to address a type of wrong she calls a “structural injustice.” A structural injustice occurs when members participating in a scheme of social coordination act blamelessly, but the scheme systematically prevents some from fulfilling their basic human rights. Because participants in a structural injustice act blamelessly, Young argues
that it is not right to hold them at fault for the resulting injustice. Although participants in a structural injustice may be blameless, Young contends that all of them acquire a form of forward-looking social responsibility to remediate the resulting harm by organizing, voting, protesting and pressuring institutions.

Young defends her theory of responsibility by arguing that traditional individualist theories of responsibility (both forward-looking and backward-looking) fail to adequately address structural injustices. If genuine structural injustices founded primarily on blameless participation exist, then traditional individualist theories of responsibility, such as an authorization account, may prove inadequate. But I argue that Young has not provided us any reason to think that structural injustices founded primarily on blameless participation actually exist. Genuine examples of structural injustices entail substantial blameworthy participation. Purported structural injustices that do not involve blameworthy participation amount to broader social or political failures that ought to be addressed with traditional duties of beneficence or distributive justice. So pace Young, I argue that we do not need to go beyond traditional individualist theories of responsibility (forward-looking or backward-looking) to adequately address structural injustices.

In the next chapter, I consider Larry May's (1992) social existentialist theory of assigning responsibility to group members. Larry May proposes that any individual who cultivates hateful attitudes shares responsibility for the wrongdoing of her fellow community members who also cultivate similar attitudes. For instance, May argues that all people in a community who bear racist attitudes share responsibility for hate crimes carried out by other members of the community. Even members who have not clearly expressed their hateful attitudes share in this responsibility. In response, I argue that May's attempt to extend shared or collective
responsibility to racist group members establishes some dangerous moral precedents. He risks trivializing the concept of moral responsibility, elides important distinctions between negligence and intentional action, unfairly blames some for actions that have no clear connection to what they did, assigns disproportionate magnitudes of responsibility and creates worrisome line-drawing problems. For these reasons, we ought to reject his account and stick with an authorization account of distributing collective responsibility.

The clearest institutional impact of a theory of distributing collective responsibility is the extent to which it does or does not justify criminal sanctions for participation in collective wrongdoing. Having defended an authorization theory of distributing collective responsibility, in the final chapter I demonstrate its relevance to international criminal law. International courts have so far introduced a number of legal doctrines to connect individual perpetrators to collective crimes: the doctrine of command responsibility, the doctrine of “joint criminal enterprise” and the International Criminal Court’s “control” standard. While each doctrine has its virtues, each also raises normative concerns. The doctrine of command responsibility unfairly attributes the crimes of subordinates to commanders who were merely negligent, joint criminal enterprise attributes global collective crimes to marginal participants and the ICC’s control standard rests on a contestable and unclear concept of control. I apply an authorization theory of distributing responsibility to show where each doctrine succeeds and where each one goes astray. I conclude by offering suggestions for modifying the doctrines to bring them onto a firm normative ground.

H. Gomperz (1939) writes:

Individual responsibility is not the only known form of responsibility; to consider it as such is characteristic of an individualistic age; in earlier stages of civilization archaic forms of collective responsibility were held to be even more significant; and in future stages some form of social responsibility seems likely, to a considerable degree, to supersede it. (p. 332)
Gomperz argues that our highly complex and socially connected age calls for a renewed look at theories of collective responsibility. He suggests that the need for collective responsibility requires us to reevaluate the individualistic principles of the Western, liberal moral tradition. Many contemporary theorists make similar proposals. The authors I cover in this dissertation either argue that we ought to weaken the strong moral link between control over the production of events and blameworthiness for their outcomes (Kutz, 2000, pp. 116-7; May, 1992) or that we need a novel theory of social responsibility based on implication or shared attitudes to distribute fault or forward-looking responsibilities.

It is my contention that these suggestions are misguided. An authorization account of distributing collective moral responsibility advances beyond the confines of H.D. Lewis’s (1948) narrow individualism while successfully explaining how individuals become responsible for the conduct of their fellow group members. While an authorization account of distributing collective moral responsibility widens the boundaries of moral responsibility, it does so without violating core liberal and individualistic moral norms of proportionality and the moral separateness of agents. Since there exists a workable theory of collective moral responsibility consistent with liberal moral principles of fairness, we should not be so quick to discard those principles in the pursuit of more expansive theories of distributive collective responsibility.
CHAPTER ONE

COLLECTIVE RESPONSIBILITY AND TACIT AUTHORIZATION

Children learn about the power of cooperation from a young age. The destruction a single child may inflict on her parents’ home and possessions is somewhat limited, for she can only reach so far and throw so hard. She will find it hard to escape the watchful eye of her parents. But once the child has friends and siblings, things become different. Now, she has accomplices: someone to steady the chair so she can reach the upper cabinets, someone to help lift the dog onto the kitchen table, someone who can keep watch to make sure mom is still outside on the patio. Her individual chaos becomes collective.

The parental response to such collective mischief is often indiscriminate. When dad walks in and discovers that his children have spread his books across the kitchen floor and covered them with flour, he will likely blame all the participating children for the mess. The moment they are found out, the children become disciples of the great individualist H.D. Lewis. They say: “Sure, I helped. But it is not all my fault.”

The child who utters this phrase has stumbled onto a profound moral dilemma. Is the collective destruction actually all her fault, or is she merely responsible for her role in the sibling conspiracy? And what about the children who did not actively participate in the mischief, but merely stood by laughing, watching or encouraging? We receive no clear answer by looking at how courts deal with collective wrongdoing. U.S. courts typically assess similar sanctions for accomplices and principals to a crime, but judges are often given personal discretion in
sentencing for a variety of reasons (Doyle, 2014, pp. 4-7). Charges of accomplice liability are far more controversial in European courts (Petersson, 2013, p. 863). International courts have developed various doctrines of collective criminal liability, from joint criminal enterprise to command responsibility, but many worry about their legitimacy (Ainley, p. 415; Darcy, 2007a, pp. 359-66; 2007b, pp. 397-9; Osiel, 2005a, pp. 798-800; Schabas, 2003, pp. 1034-6). It turns out that determining the correct magnitude of blame for collective wrongdoing is as difficult for adults as it is for children; sometimes it seems acceptable to blame one person for the actions of another, while sometimes it does not.

I aim to ease this confusion by developing a workable theory of distributive collective responsibility. I propose we understand collective responsibility as a function of authorization. Usually, an individual acts wrongly by intentionally transgressing a moral norm and we hold her to blame accordingly. However, an individual may also authorize another to act on her behalf. If one individual A successfully authorizes another individual B to do wrong on her behalf, A is to blame for the action or actions that B does on her account. Because A willingly makes B an extension of her autonomous capacities, it is fair to hold her responsible for B’s wrongdoing.

One could grant that individuals are to blame for the actions they expressly authorize other agents to perform, but such a stipulation would not amount to an interesting account of distributive collective responsibility. Examples of express authorization do not seem morally problematic. When individual A gives individual B power of attorney, we think it right for B to make decisions on A’s behalf. If A authorizes B to carry out some legal but ethically dubious course of actions, we do not hesitate to blame A for B’s actions (Pogge, 2008, pp. 83-6).
Similarly, if a husband contracts a hitman to kill his wife, few people find it a stretch to blame the husband (as well as the hitman) for his wife’s murder.

The problem is that collective action is usually not coordinated with express agreements of authorization. When a group of bullies get together to beat up a middle schooler, one bully does not say to another: “Sean, I expressly authorize you to punch Herbert in the stomach on my behalf.” A theory of distributing collective responsibility based solely on express authorization simply will not extend to cover most instances of collective wrongdoing. A theory of distributing collective responsibility should give us more.

We therefore need to augment a theory of express authorization with an account of tacit authorization. Sometimes one expressly authorizes another to act on one’s behalf, but other times one tacitly or implicitly authorizes another to act on one’s behalf through one’s active participation in some venture. By intentionally making oneself a member of a common plan and contributing to the realization of the common plan, I argue that one tacitly or implicitly authorizes one’s fellow group members to help realize the collective end on one’s behalf. An agreement of tacit authorization is therefore an actual agreement. However, it is an agreement that is not made through formally stipulated means of expression. Tacit authorization usually functions in cases where an express agreement would be “otiose or redundant” (Carr, 1990, p. 338).

In this dissertation, I argue that an authorization theory of distributing collective responsibility is the theory of distributive collective responsibility that adheres most firmly to
basic principles of fairness.\(^1\) Because this is so, proponents of an authorization theory of distributing collective responsibility can successfully respond to skeptics of collective responsibility who claim that the concept of collective responsibility is intrinsically unfair (H.D. Lewis, 1948; Narveson, 2002, p. 190; Velasquez, 1983; 2003). As I argue in later chapters, some major competing theories of distributive collective responsibility struggle to answer charges of unfairness. Because an authorization theory of collective responsibility successfully defeats worries of unfairness, theorists have a normative reason to prefer an authorization theory over its peers.

To frame my argument, I begin with a brief discussion of two concepts that are integral to my project: responsibility and authorization. Both concepts have different shades of meaning in the literature, so it will be important to begin with a precise and accurate understanding of both. I then proceed to lay out two principles of fairness relevant to ascriptions of responsibility: the separateness of agents and proportionality. I will use these principles as a normative standard to assess the fairness of any theory of distributive collective responsibility. I proceed to articulate an authorization theory of distributing collective moral responsibility. I argue that if an agent freely and intentionally joins a common plan, has accurate beliefs about the common plan and makes a substantial contribution to the common plan, that agent tacitly authorizes the other members of the common plan to act on her behalf. I then demonstrate that an authorization theory of distributing collective responsibility is in accord with both principles of fairness.

\(^1\) In some philosophical discussions, the concept of fairness is interpreted very narrowly to mean an equitable distribution of goods; see Rawls (1971). I use the word fairness here in a much broader sense, to cover not only the proper distribution of goods, but also the equitable distribution and assignment of fault and liability.
Responsibility and Authorization

It is of no use to advance a theory of distributing collective responsibility without nailing down a definition of responsibility first. Theorists speak of responsibility in different ways, and this has led to confusion. One can broadly categorize responsibility as either forward-looking or backward-looking (Young, 2011, p. 96). One form of backward-looking responsibility is causal responsibility. One invokes causal responsibility with such phrases as “the hurricane is responsible for this mess” and “the pressure change is responsible for my headache.” Causal responsibility makes no reference to free will or intentionality. To say that A is causally responsible for B is merely to say that A caused B, however one wishes to define the causality relationship.

Causal responsibility differs from backward-looking moral responsibility. Backward-looking moral responsibility attributes a state of affairs to an agent and assesses the agent’s performance as negative or positive. A designation of being “to blame” for a state of affairs negatively credits the action to an agent’s record or reputation (Feinberg, 1970, p. 128). When a performance is positive, the agent is commended or credited for it. Frequently, when a person is to blame for a state of affairs, the person also merits or deserves some sort of punitive reaction. Similarly, when a person is credited or commended for a good action, we tend to think she merits praise or reward. But this is not always so. One can be to blame for a state of affairs yet not liable to receive any sort of punitive reaction. This happens most frequently when one causes a freak accident or intentionally performs an action due to an unreasonable level of coercion. I will therefore follow Feinberg (1970) in separating negative assessments of backward-looking moral responsibility into being at fault (or to blame) and being liable. The former attributes a defective
performance to an agent while the latter ascribes a normative property to the agent, that of meriting or deserving some reaction.

It is worth saying a little more about liability. The most common informal, non-institutional punitive reaction in the literature is one of blame (Gilbert, 2006, p. 97). To say that an agent is morally liable for a state of affairs is to say that one is justified in blaming the agent for that state of affairs. This means that others may justifiably adopt attitudes of anger and resentment toward the agent and rightfully hold the belief that the agent was in the wrong for her actions. In some cases, blame may be too harsh a reaction. Some theorists argue that moral liability sometimes calls for a person to feel shame, even if it is not appropriate for others to upbraid her in anger (May, 1992, p. 120). Ascriptions of backward-looking liability are normative. They do not stipulate that a person is or will be blamed for a state of affairs. They stipulate that a person should be blamed for the state of affairs and that blaming the person for the state of affairs is justified.

As for basic ascriptions of fault, one might be to blame (or at fault) for an action or the consequences of an action (Sverdlik, 1987, p. 64). One can be to blame for killing or murdering another or one can be to blame for her death, however it may have come about. It is more controversial among moral philosophers, but one might also be to blame for the state of one’s character (Feinberg, 1970, p. 191; May, 1992, 52-4; Zimmerman, 2002, p. 563). If one’s character is the result of many intentional decisions, then one can be credited or discredited for the development of one’s dispositions and habits. A person who has spent his lifetime indulging his desires to gamble and drink may be to blame for his weak will, just as a star athlete may be credited for developing her capacities through hard work.
Backward-looking ascriptions of responsibility differ from forward-looking ascriptions of responsibility. Forward-looking responsibility ascribes a task or role that an individual ought to perform in the present or future (D. Miller, 2001; Stilz, 2011, p. 195; Young, 2011 p. 96). To say that BP is responsible for cleaning up the oil spill in the Gulf Coast is to say that the corporation has an obligation to take certain measures to remedy the harm. Theorists sometimes refer to this sort of forward-looking responsibility as “remedial responsibility,” the responsibility to make a bad situation right (D. Miller, 2001, p. 454). However, forward-looking responsibility need not be remedial. Forward-looking responsibility sometimes involves an obligation to ensure that a certain state of affairs obtains (Smiley, 2014, p. 2 n.1). For instance, to say that a parent is responsible for the education of her child is to say that the parent has an obligation to ensure that the child is properly instructed.

It is important to keep causal, backward-looking and forward-looking responsibility separate. It may be the case that one can be to blame for a certain event, even if one was not an apparent causal factor in bringing it about. Consider Homer Simpson, an employee of the Springfield power plant. If Homer Simpson falls asleep on the job and the Springfield reactor melts down, Homer is to blame for the meltdown, but not because he stands in an obvious causal relationship to the catastrophe. Homer did nothing, and that is precisely the problem. Homer is to blame because he should have acted to prevent the meltdown. So sometimes individuals may be to blame for not causing a state of affairs to obtain.²

² Hart and Honoré (1985) argue that omissions have causal efficacy, while others contest this point; see Dowe (2001) and McGrath (2005).
Similarly, while assignments of remedial responsibilities usually relate to assignments of fault, this is not always the case. Sometimes we assign remedial responsibilities to some for the faulty actions of others. For instance, in American civil law, when an individual acting as an employee of a corporation wrongs another, the harmed individual may sue the corporation for damages (Davant, 2002, p. 511). Even though the individual who commits the wrongful act is to blame for her intentional actions, her company and superiors are held responsible to remedy the harm. This is more common for non-remedial forms of forward-looking responsibility. Parents are responsible for the upbringing of a child not because they have done something wrong or faulty, but because they have a special relationship with their child and are most suitable for ensuring her proper development.

When I speak of moral responsibility throughout this project, I typically have backward-looking fault in mind. The concept of fault is often crucial to discussions of collective responsibility (Sverdlik, 1987, p. 62). I will designate fault-responsibility as moral responsibility with a subscript f. One specific reason I focus on backward-looking fault is because I wish to develop a theory of collective responsibility that will either morally justify or not justify legal punitive measures for collective wrongdoing (see Corlett, 2001, ch. 8).

Assuming even a weakly retributive or expressive theory of punishment, it is plausible to stipulate that if one is to blame for an action or outcome, then one is prima facie worthy of being punished for that action or outcome (Feinberg, 1970, p. 218). One is only prima facie or pro

\[^3\text{It is worth noting that most theorists run together what I, following Feinberg (1970), have designated “fault” and “liability.”}\]
tant to liable to punishment because institutional punishment should take into account other practical considerations besides an agent’s fault (Feinberg, 1970, p. 40).

Before moving on to a discussion of authorization, I will address a potential area of concern that is popular in the contemporary literature on moral responsibility\(_f\). This concern involves the importance of the connection between free will and moral responsibility\(_f\). Incompatibilists argue that moral responsibility\(_f\) requires the existence of a metaphysically free will, while compatibilists argue that moral responsibility\(_f\) could exist even if the universe turns out to be a closed, physical system (Fischer and Ravizza, 1993; Van Inwagen, 1978; 1999). For the most part, I have little stake in this discussion. An authorization theory of distributing collective moral responsibility\(_f\) is consistent with most forms of compatibilism and incompatibilism. I say “most” because there is an exception. Recently, some theorists have argued that incompatibilism is true and that free will does not exist. If moral responsibility\(_f\) requires free will and free will does not exist, the argument goes, we should stop holding people responsible\(_f\) for their wrongdoing (see Pereboom, 2001). Were this true, any theory of collective moral responsibility\(_f\) would be suspect.

If incompatibilism and determinism turn out to both be true, this would be damaging to my project. It would be harmful to any project on moral responsibility\(_f\). But physicists have yet to prove outright that determinism is true and moral theorists are far from producing a decisive argument in favor of incompatibilism. I also contend that even if theorists discredit moral responsibility\(_f\) as a metaphysical thesis, an authorization account of distributing moral responsibility\(_f\) could survive as a moral thesis. If utilitarianism is correct and a policy of holding people responsible\(_f\) for those wrongs they authorize will likely produce the best consequences of
all policy options, then an authorization account of collective moral responsibility would be morally viable. Or perhaps an authorization theory of collective moral responsibility would be the option preferred by all rational agents choosing schemes of social cooperation from some idealized position (Rawls, 1971, pp. 17-22; Scanlon, 1998, pp. 202-6). This could confer legitimacy on a policy of distributing collective responsibility whether or not free will exists. I do not explore either possibility here, but they would both be live options even if incompatibilism and determinism are shown to be true. I will therefore leave the battle over skepticism to the physicists and metaphysicians and proceed.

Let us now turn to authorization. There is not a great deal of analysis of the concept of authorization (Volmert, 2012, p. 287). Discussions of authorization in the philosophical literature generally relate to the problem of democratic authorization (Hobbes, 1985, ch. 16; Pitkin, 1967, ch. 3; Estlund 2008, p. 65). In these discussions, authorization functions as a justification for sovereign rule. According to an authorization theory of representation, a sovereign justifiably makes decisions for the state only if the sovereign has been authorized by the electorate (Pitkin, 1967, p. 38). The electorate authorizes the sovereign to rule by voting and participating in the democratic process.

While many theorists of democracy and representation utilize the concept of authorization, few give a clear account of its meaning (Volmert, 2012, p. 287). Volmert (2012, p. 288) attempts a definition, so I begin there. He suggests we define authorization as a practice that

---

4 For one such “normative” account of moral responsibility grounded in the social contract, see Lenman (2006). See also Wallace (1994).

5 A good deal of what has been written about authorization comes from discussions of Hobbes’s political theory; see Copp (1980) and Gauthier (1969, ch. 4).
“(1) involves giving someone permission to do something, and (2) presupposes that the
authorizing party has a right to decide what happens within the domain at state.” As Volmert
notes, these criteria are not exhaustive of our concept of authorization (p. 288); they are
necessary but not sufficient for an authorization relationship. These two criteria must be
supported by social norms: it must be the case that members of a community recognize the
binding force of agreements of authorization (p. 288).

I have an initial concern with Volmert’s (2012) definition. Volmert’s definition makes no
clear distinction between authorization and the practice of giving consent\textsubscript{p}.\textsuperscript{6} If A consents\textsubscript{p} to sex
with B, then B has permission to engage in physical contact with A that is usually impermissible.
B can give this consent\textsubscript{p} because B has a right to decide what happens to her body. There seems
to me to be a crucial difference between such giving of permissive consent\textsubscript{p} and authorizing
another to act on one’s behalf.\textsuperscript{7} When I authorize another to negotiate a contract on my behalf, I
do not necessarily permit an infringement of my rights. Rather I enter into an agreement such
that the other person may further my interests and bind me to new obligations.

It is common to use the word “consent” to denote both acts of permitting and acts of
authorizing. The two concepts are similar, but it is important to keep them separate. Some
agreements involve both authorization and consent\textsubscript{p}, but some agreements involve only one or

\textsuperscript{6} By consent I mean only what some refer to as “permissive consent,” the consent one gives to permit another to
infringe upon one’s rights, interests or personal sovereignty; see Manson (2016). I will use the subscript \textsubscript{p} to denote
permissive consent, as opposed to the broader notion of consent which may also include such acts as promising to
do one’s part in a cooperative plan.

\textsuperscript{7} I do not deny that sometimes an individual may give a form of permissive authorization to allow others to advance
their own interests by accessing or utilizing that individual’s property interests, as when a university authorizes a
student group to access the gym for basketball practice. For the sake of this dissertation, I am not so interested in
permissive forms of authorization, as it is unclear whether giving permissive authorization to others entails any
meaningful transfer of moral responsibility on account of what they do with that authorization; see Atenasio (2018,
n. 19).
Consider a patient that consents to an operation. The patient thereby permits the doctor to cut her open. However, the operation is done not for the sake of the doctor, but for the patient’s sake. While it may not be commonly spoken of this way, it is possible to stipulate that a patient also authorizes the doctor to carry out the operation. In performing the operation, the doctor acts on the patient’s behalf. This would be stranger to say of one person having sex with another. It is certainly possible for A to authorize B to have sex with A on A’s behalf. However, most sexual acts do not involve this sort of authorization; they are between two willing parties who give their permission to one another. Individuals who have sex are not necessarily acting on each other’s behalf.

Keeping in mind the similarities (and differences) between consent and authorization, I develop a definition of authorization more specific than the one Volmert (2012) provides. Definitions of consent are common in the literature, so I will adapt a theory of authorization from a contemporary theory of consent. Following Malm (1996) I define consent as “the signification of a particular mental state through the performance of a conventionally recognized act (or set of acts, or one of a set of acts)” (pp. 147-8). This is generally known as a performative theory of consent. Performative theories of consent stipulate that consent requires the performance of a specific action in a certain context. They are distinguished from mentalist theories of consent, which stipulate that the giving of consent involves instantiating a particular mental state (Alexander, 1996; Hurd, 1996).

I prefer a performative account for the reasons given in Malm (1996) and Dougherty (2015), but it is possible to develop a theory of distributing collective responsibility based on tacit authorization utilizing a mentalist theory of authorization as well. I will say more about the advantages and disadvantages of a mentalist theory of authorization later in this chapter.
Some argue that the relevant mental state is one of desire. However, the signification of desire or want is not quite the same as giving consent\textsubscript{p} (Malm, 1996, p. 148). One can consent\textsubscript{p} to something while strongly hoping and desiring that it not occur. A person can also consent\textsubscript{p} to sex with another while simultaneously indicating to the other how little she desires it. For these reasons, I suggest we follow Malm (1996) in designating the mental state as a state of “willingness” in the sense of “willing to go along with” (p. 148). The mode of one’s expression of willingness must be one that, given one’s social context, plausibly signifies one’s willingness to others.

Much the same can be said of authorization, but it differs from consent\textsubscript{p} in an important way. Whereas consent\textsubscript{p} signifies a willingness to permit an infringement of one’s rights and personal sovereignty, authorization signifies a willingness to have someone carry out an action as one’s proxy agent. In consent\textsubscript{p}, I demonstrate to another party that he or she is permitted to do something to me that, under normal circumstances, would be impermissible. In authorizing, I demonstrate to another party that he or she is permitted (or obligated) to carry out an action for me, and that in performing the action, the other party is acting as an extension of my own autonomous capacities.

I thus stipulate the following: A authorizes B to act on A’s behalf only if A signals to B A’s willingness (in the proper social context) to have B carry out an action or set of actions for A. The actions in question may be specific directives, such as when one person authorizes another to write a will, or they may be quite general orders, such as the directives an individual gives when she authorizes another to invest her capital wisely.
When A authorizes B to carry out an action on A’s behalf, A and B enter into something akin to what Anglo-American law denotes as an agency relationship. B becomes an agent of A. However, it is important to avoid an overly simplistic interpretation of agency. One way to interpret agency law is that, when A enters into an agency relationship with B, A uses B “merely as an instrument” for her own purposes (Kadish, 1985, p. 370). On this interpretation, because the agent is more or less a tool for the principal, when the agent commits wrongdoing, it is as if the principal commits the wrongdoing herself.

This is too rudimentary an understanding of agency. You can use another as an instrument without forming any sort of agreement. Consider a thief who wants to gain access to a wealthy investor’s private vault. The vault only opens with a scan of the investor’s thumb. The thief hides near the vault, and when the investor arrives, the thief knocks the investor unconscious. The thief then takes the investor’s limp hand and presses the investor’s thumb to the scanner, thereby gaining access to the vault. In this case, the thief has used the investor’s body as an instrument for her purposes, but formed no voluntary agreement. For this reason, I think it best not to view the agent as a mere instrument of the principal actor. An act of authorization does not efface the personal or moral agency of the agent. The agent continues to act on her own volition. Rather, when B freely and intentionally acts on the authorization of A, they effectively commit the action together. There is a “concurrence” of wills, to use a medieval expression. If A authorizes B to carry out an action on A’s behalf, it is if A and B jointly and intentionally carry out the action together.

While so much suffices for a basic account of authorization, there is still a crucial distinction to discuss. The account of distributing collective moral responsibility, I propose relies
on a specific form of authorization: tacit authorization. There is no literature on the concept of tacit authorization, so for a discussion of the concept, it makes sense to begin with a look at theories of tacit consent. I first consider Simmons’s (1979) definition of tacit consent, as it is popular in the literature. He argues that tacit consent is consent given by “remaining silent and inactive” (p. 80). For Simmons, the difference between tacit consent and express consent is whether an agent expresses her willingness through intentionally acting or intentionally remaining inactive. The two ways in which a committee chairperson might hold a vote bring out this difference. A committee chairperson might say to the committee “all in favor, raise your hand.” In this case, those committee members who raise their hands would be expressly consenting to the policy. Conversely, the committee chairperson might say “any objections to this policy?” If the committee members remain silent and “fail” to raise their hands, then they have tacitly consented to the policy (Simmons, 1979, pp. 79-80).

For Simmons, the difference between tacit and express consent is in the mode of expression. He argues that tacit and express consent are both fully binding forms of social agreement (p. 80). The difference is that properly expressing one’s intentions through words and deeds counts as express consent, while properly expressing one’s intentions through silence and inaction counts as tacit consent (p. 77). In both cases, successful consent requires some background conditions to obtain. Conditions must be in place such that one’s social group understands the relevant actions, words or silence to constitute consent (pp. 80-81).

I think Simmons is on the right track here, but I would like to clarify, or perhaps modify, his understanding of tacit consent (see also Harris, 1992, p. 664). Consider the voting example

---

9 Plamenatz gives a similar definition in (1968, p. 8).
again. Let’s say I am a voting member of a large social organization. The chairperson calls a vote and asks all in favor to raise their hands. At the same moment, I happen to be stretching my arms upward and yawning because I am tired. However, I do not want to vote in favor, so when the chairperson calls the vote, I quickly lower my hand. In this example, have I voted tacitly or expressly? On the one hand, I voted by making sure not to raise my hand. However, in order to vote my mind, I had to act by lowering my hand to get my body into the proper position so that my inactive position would successfully signal my intentions to the chairperson.

The danger here is that we misconstrue what it means to remain silent and inactive. One way to interpret “silent and inactive” is to mean no more than trying one’s best not to move or speak. On this interpretation, tacit consent\(_p\) would result when one consents\(_p\) by being quiet and sitting still. If we understand tacit consent in this way, then when I lower my hand to vote, I vote expressly by moving my body to a neutral position. Had I not been stretching, I would have voted tacitly, by remaining still in my seat and not moving my body. The difference in tacit and express consent\(_p\) would then turn on the starting point of my body when the chairperson calls the vote. This makes the distinction seem rather trivial or irrelevant.

Perhaps a better way to understand “silent and inactive” is as relative only to a set of formally recognized words or gestures. To give tacit consent\(_p\), I need only signal my willingness to others without using a specified word, statement, gesture or action. I remain silent because I do not utter the actual words “I consent to this,” and I remain inactive insofar as I do not perform a specified action, such as raising my arm to vote. However, it is important to note that there are other ways to signal one’s willingness tacitly besides sitting still and refraining from all language use.
Consider a clinic that gives out free flu shots in a poor neighborhood. The lines for shots are very long, so the nurses set up an expedited line where no paperwork is necessary (ignore for a moment the legal implications of such a practice). Each patient sits down at the expedited station and gets a shot in the arm. Now, stabbing someone in the arm is usually morally impermissible. The shot does not become permissible because of any patient’s express consent, as none have spoken, nodded or made any conventionally recognized sign of assent. Neither are the patients sitting perfectly still. In order to receive the shot, the patients must sit in the chair, roll up their sleeves and present their arms. They signal their willingness to receive the shot by going along with the social practice.

Therefore, instead of using the language of “silent and inactive,” I think it is clearer to understand tacit consent as consent given by participating in or going along with a venture to the point that one’s behavior becomes sufficient to signal one’s willingness to others, depending on the background conditions of one’s social context. Sometimes, one goes along with a venture by sitting still or remaining quiet to avoid dissenting. Other times, one goes along with a venture by other means of participation.

I therefore suggest the following as the difference between tacit and express consent. Express consent occurs when one expresses one’s willingness by making a move in what Wittgenstein calls a language game (1953, §19-25). This “move” could be a statement or recognized gesture meant to stand in for a statement. Tacit consent occurs when one expresses one’s willingness through one’s participation in some endeavor. Tacit consent is not given by a single, conventionally recognized statement or gesture. Rather, it is given by actively going along with an activity. One’s participation in or going along with a venture constitutes tacit
consent, when one’s behavior reaches a threshold that, given one’s social context, plausibly signals one’s willingness to be a part of the social practice or venture. If, for example, a person attacked the nurse in retaliation for being stuck with a needle, we would consider her actions unjust. The person waited in line and had ample time to understand what it meant to sit in the chair. She has no justifiable grievance against the nurse, because by waiting in line and sitting in the chair, her behavior is more than sufficient to signal her willingness to receive the shot. In other words, she has given her tacit consent to be stuck with the needle.

One might worry with Simmons (1979, p. 88) that tacit consent (or authorization), as I have defined it, is not a sign of consent but merely something that “implies” consent. He goes on to argue that implying consent is not the same thing as actually giving consent, and that it therefore has no moral force. But this seems to be a concern not with the nature of consent, but with the epistemology of recognizing consent. All acts of consent imply something to nearby observers: those acts imply that the individual expresses her willingness knowingly and intentionally. Any individual who wishes to determine if another has consented must infer the other person’s willingness from her actions. She either infers the other’s willingness from some explicit statement or gesture or from her participation in some activity. Inferences can be correct or incorrect in cases of tacit and express consent. If I vote in favor of a measure because I am stretching, while the chairperson may infer my willingness, I have nonetheless not consented intentionally. My gesture was an accident. One can raise skeptical concerns about the validity of express consent, just as about the validity of tacit consent. The concern about actions “implying consent” is therefore a skeptical problem with identifying consent in general, not tacit consent specifically.
One might also worry that, given my definition of tacit consent, one can tacitly consent to a venture or social practice while intending, wishing or hoping that it not occur. This would seem to create a practical paradox, for how can one tacitly consent to something one would never expressly consent to? Imagine that an individual comes upon one of the nurses doling out free flu shots. The individual stands in line for 15 minutes, sits in the chair, rolls up her sleeve and allows the nurse to stick her with the needle. As soon as the needle is withdrawn, the individual objects: “when you stuck me with the needle, I was actually unwilling to have you do that. Despite my apparent behavior, my inner will was decidedly against the shot.” If the individual’s story is plausible, then this is a problem for an expressive or performative theory of tacit consent. It is logically impossible to both consent to something and not consent to it at the same time and in the same respect. By tacitly consenting to the shot while expressly not consenting to it at the same time, this seems to be what our individual is doing.

But this objection rests on a faulty theory of action. We can see why by examining a clearer case of express consent. Imagine instead that the individual in question must sign a form giving her consent to receive the flu shot. The individual signs her name on the form and then sits in the chair, rolls up her sleeve and receives the shot. The individual then claims: “while I was signing the form, my inner will was completely against receiving the shot.” A problem arises when we try detach the idea of an inner will from intentional performances that embody that will. It is frankly absurd to say that you can freely and intentionally carry out an action while simultaneously willing that it not occur. You may regret or dislike having to perform an action, but I can make no sense of performing an action intentionally while willing its
nonperformance. If a criminal tells a judge: “I did strangle the victim to death intentionally, but my inner will was against the action the whole time,” the judge will rightfully reject this defense as utterly nonsensical, and justifiably so.

If one freely and intentionally expresses one’s willingness to receive a flu shot by signing a consent\textsubscript{p} form, then one is willing to receive the shot. If one freely and intentionally expresses one’s willingness to receive the flu shot by standing in line, sitting down and rolling up one’s sleeve, then one is willing to receive the shot. Now, it is possible for a person to pretend to express her willingness to go along with some activity. Perhaps the individual standing in line to receive the flu shot is a government agent who is spying on the free clinic, but who has no interest in receiving the shot. But this sort of false consent\textsubscript{p} is a problem equally for express and tacit consent\textsubscript{p}. I can pretend to express my willingness by disingenuously signing a form or I can pretend to express my willingness by appearing to cooperate with a social practice. In both cases, we can argue about whether consent\textsubscript{p} actually occurs or not. However, whether or not pretend or false consent\textsubscript{p} constitutes actual consent\textsubscript{p} is a different discussion than whether tacit consent\textsubscript{p} is a valid form of consent\textsubscript{p} or not.

Assuming that these objections have been sufficiently addressed, we can now translate this definition of tacit consent\textsubscript{p} into a definition of tacit authorization. If tacit consent\textsubscript{p} occurs when one expresses one’s willingness to relinquish a personal right or aspect of one’s personal sovereignty by sufficiently going along with a joint venture, then tacit authorization occurs when

---

10 The astute reader here will detect the influence of Wittgenstein (1953) and Anscombe (1957) in my rejection of a strong distinction between will and bodily action. But note here, even Descartes rejects gnostic attempts to separate the will from the body too sharply. In the Sixth Meditation, he writes that “I am not merely present in my body as a sailor is present in a ship, but... I am very closely joined and, as it were, intermingled with it, so that I and the body form a unit” (1984, p. 56). In The Passions of the Soul, Descartes writes that “our merely willing to walk has the consequence that our legs move and we walk” (1985, p. 28).
one expresses one’s willingness to have another act as one’s proxy agent by sufficiently participating in or going along with a joint venture. The level of sufficiency is determined by when, given the nature of the action and the context in which it occurs, it would be plausible for others to believe that one has sufficiently expressed one’s willingness to constitute an authorization relationship.

**Tacit Authorization**

Tacit authorization occurs when one’s activities and participation in a group activity or common plan reach a threshold at which it becomes plausible to interpret one’s actions as a sufficient expression of willingness to have the other group or common plan members act on one’s behalf. The next step in the argument is to define approximately where this threshold lies. In other words, we need to determine the level of activity or participation in a group action or common plan that establishes an authorization relationship to the other members of the group or common plan.

I argue that the following are necessary and sufficient for the tacit authorization of fellow members of a group or common plan. First, an individual must freely and intentionally make herself a recognized member of the group or common plan. Second, an individual must be sufficiently informed and have correct beliefs about the relevant activities and aims of the group or common plan. Finally, the individual must freely and intentionally make a substantial contribution to the realization of the common plan. Any individual who meets these criteria thereby tacitly authorizes the other members of the group or common plan to perform actions on her behalf in pursuit of a shared end. The upshot is that any group member who tacitly authorizes
the other members to bring about a shared end is to blame for the other group members’ (relevant) wrongful actions and the reasonably foreseeable consequences of those actions.

The first criterion for tacit authorization I propose is a group or common plan membership stipulation. To tacitly authorize one’s fellow group members to act on one’s behalf, one must first of all be a recognized member of the group or common plan. If an individual is a member of a group or common plan, then one or more persons will recognize her (or have the potential to recognize her) as someone who fits into some scheme of social relation or cooperation. In some organizations, membership may entail strong normative expectations. The president or CFO of a corporation has obligations pertaining to the company’s daily operations. Other membership requirements come with far weaker or no expectations. One could become a registered member of an online discussion group without acquiring any sort of obligation to participate.

Recognition is an integral component of group membership. Recognition by one’s fellow group members facilitates coordination and the delegation of tasks and opens up new avenues for cooperation. However, group membership does not necessarily require one group member to actively recognize another. Member recognition may be no more than possible or dispositional in one or more group members. It need only be the case that, were group member A to utilize the relevant resources to confirm another person B’s membership, A could in principle successfully do so.

It is possible to object here that someone may attempt to join a group while remaining unrecognized as a group member. Consider the example of a secret agent. The agent wants to infiltrate a nationalist organization to spy on their activities. The agent therefore desires to
become a member of the nationalist organization while avoiding recognition. But if the agent is able to become a group member while avoiding recognition, then it would turn out that recognition is not a crucial component of group membership.

I do not think this example defeats the need for recognition in group membership. The secret agent may infiltrate the group in one of two ways. She could remain completely invisible to the nationalist organization. She could infiltrate their listserv, listen to their phone conversations and photograph their meetings. But if the secret agent takes this route, she is not plausibly a member of the group. It is unclear to me why anyone would call her a member of the group if the other group members have no knowledge of her existence.

The secret agent could also put on a disguise and join the nationalist organization under false pretenses. She could claim to share the nationalists’ love of blood and soil. She could get herself invited to the meetings and attempt to rise into the ranks of leadership. But if the agent takes this path, then she has made herself a recognized member of the group. At least some nationalists in the group could recognize her false persona as a group member. Now, they do not recognize the secret agent’s true identity, but that is not the notion of recognition that interests me. When I say that membership requires recognition, I only mean recognition as a group member. Others must be able to confirm that one is, indeed, a member of the group or common plan. If recognition among one’s fellow group members is not possible (as in the example of the agent who remains invisible) then one cannot count as a member of the group. Either the secret agent remains invisible and she is not a group member or she joins the group under false pretenses and counts as a member of the group (albeit one that is working against the interests of the other group members). Either way, recognition of some sort tracks group membership.
To tacitly give one’s authorization to one’s fellow group members, it is not enough that one become a member of a group or common plan, one must do so intentionally. Imagine that the aforementioned nationalist group begins signing all “pure blooded” citizens up for membership. Anyone with an acceptable last name is automatically enrolled. In this case, while the conscripted individuals are recognized as group members, they did not become members intentionally. Rather, others sign them up without their approval or knowledge. When we discuss what the nationalist group does collectively, it makes no sense to speak of people who have been included in the group without their agreement or approval.

It is not sufficient that one join a group intentionally, one must also join the group freely or voluntarily. There are cases in which an individual signs up for a group intentionally, but does so only because she is forced or coerced. Consider the example of a police officer pressured by the members of a mafia family. Imagine that the mafia family takes the police officer’s son hostage. Members of the mafia family then tell the police officer that unless the police officer joins the family’s scheme and helps further their criminal ends, they will kill the officer’s son. The police officer reluctantly agrees to be a part of the mafia family’s criminal plans, on the condition that the officer’s son remains unharmed. In this case, the mafia family recognizes the officer as a member of their plan, but the officer has only joined due to an unreasonable threat of force. Since we are interested in a theory of tacit authorization, it would make little sense to stipulate that by being forced or coerced into a group, one thereby tacitly authorizes one’s fellow group members to pursue some end. Tacit authorization requires an expression of willingness in the proper context, and that context certainly requires that one’s willingness be an expression of genuine (unforced) willingness.
However, there may be some valid group members to a common plan who did not join intentionally. Some people are enrolled in an organization while they are too young to make informed, rational decisions. Those individuals then embrace their role in the organization and become integral, contributing members. If one becomes a group member before one is capable of rational choice, it is sufficient to fulfill the membership stipulation that one fail to opt out of one’s membership once one is able to make autonomous choices. Assuming that one is sufficiently rational, informed about what membership entails, aware of how one relinquishes membership and capable of relinquishing membership without incurring unreasonable costs, one fulfills the membership criteria, even though one did not join the group or common plan intentionally.

There are two immediate objections to a membership criteria for distributing collective responsibility on account of tacit authorization. The first objection concerns the metaphysics of responsibility. The worry is that a requirement for group recognition introduces a component of responsibility that is external to an individual’s mental states (Ciurria, 2015; Frankfurt, 1969; Levy, 2014). The problem with introducing components external to an individual’s internal psychology is that this would introduce an unacceptable element of luck to the ascription of moral responsibility (Enoch and Marmor, 2007; Hartman, 2016; Zimmerman, 2002). Because luck is morally arbitrary, the argument goes, luck cannot play any role in the ascription of moral responsibility (Zimmerman, 2002, p. 559). If group membership introduces an element of luck into a theory of distributing collective responsibility, then group membership is an unacceptable component.
There are two possible responses to this worry. First, it is not clear that any theory of moral responsibility can entirely do away with the problem of moral luck. All agents are shaped by historical factors, such as the date and location of their birth and the conditions of their upbringing (Ciurria, p. 606). There is also a growing field of literature on social psychology that indicates that many of our actions are strongly solicited by social or external factors (Doris, 1992; 2002; Haidt, 2012, pp. 100-103). It is possible that the scope of human actions that are completely devoid of moral luck are exceedingly narrow. This would entail that most people are not responsible for that majority of what they do. Were this so, this would consign the concept of moral responsibility to triviality.

For those who contend that responsibility must be luck-free (or almost luck-free), one could modify the group membership requirement to be “in the head,” so to speak. One could stipulate that the group membership stipulation is satisfied by having an intention to become a group member, assuming one takes an intention to be some form of mental state. This step would begin to transition from a performative to a mentalist theory of authorization. Just as a mentalist theory of consent stipulates that consenting involves instantiating a specified mental state, a mentalist theory of authorization would stipulate that authorizing involves instantiating a specified mental state or set of states.

This move is possible for a theory of tacit authorization, but not necessarily preferable. One can think up examples where an individual intentionally acts to join a group, but is wildly deceived about the status of her membership. Let us imagine that there is an American who becomes radicalized by online propaganda. The American reads the writings of terrorists and decides to become a member of the terrorist organization. The American begins to post on a
secret online forum that she believes serves as a communication hub for the terrorist group. She pledges her membership to the terrorist organization on the forum, and the forum members acknowledge her membership and praise her for her courage to stand against her own country. However, the forum members are not themselves members of the terrorist group; they are a collection of teenagers who started the forum as a joke.

Let us also imagine that the American begins to send money to a known terrorist in another part of the world. The terrorist receives the money and uses it, but has no idea where it comes from. Would we want to say that the American is part of the terrorist organization and that she is an active participant in a global collective action? I’m not so sure. It may be true that the American is equally to blame as any American who succeeds in successfully joining the terrorist organization, on account of her intention to join the group. This would be the case for any luck-free, psychologically internal account of moral responsibility. But I am skeptical that luck-free moral responsibility is feasible, so I will continue to speak of the membership condition as requiring actual group recognition, not merely an intention to be a recognized member. Those who cannot tolerate moral luck or who prefer a mentalist theory of authorization are free to substitute an intention for group recognition or membership in place of intentionally joining a group.

The preceding considerations lead to the next objection, which is that membership and group recognition are not an integral component of collective action (Kutz, 2000a, ch. 5; 2000b). The only requirement for collective action is that one contribute to the realization of a shared goal or aim. The question of whether you plan on coordinating with or relying on others is
inconsequential; what is important is that your action is part of a larger set of actions related to a single goal in some way.

In ruling on the applicability of accomplice liability, American courts have historically judged that common plan or group membership is not a requirement for status as an accomplice. In State v. Tally (5 So. 722 Ala. 1894), the Supreme Court of Alabama ruled that Judge Tally could be charged with murder for his contribution to a collective action, even though the other members of the collective had no knowledge of Tally’s contribution. The judge, upon hearing that his brothers-in-law (the Skeltons) sought to murder the illicit paramour of his sister-in-law (R.C. Ross), prevented a warning telegram from making it to the paramour. The Skeltons then proceeded to murder Ross. Tally was at first found not guilty as an accomplice to murder, but the prosecution appealed the verdict. The Supreme Court of Alabama later found him guilty of aiding and abetting murder.\(^\text{11}\)

It seems that many people’s intuitions and considered judgments are against me on this point. I still wish to retain a membership component for the following reasons. First, a membership condition helps to maintain a meaningful distinction between collective intentional action and individual risk-raising, negligence or recklessness. Consider again the Tally case. Tally helped facilitate the murder, but it was the Skeltons who collectively and intentionally organized and committed the murder with no knowledge of Tally’s help. Now imagine that Tally facilitated Ross’s death by preventing a telegram that would apprise Ross of a natural disaster

\(^{11}\) The fact that Tally was not initially charged with murder should indicate that such cases of marginal aiding and abetting remain somewhat controversial. In addition, accomplice liability is often understood as a less serious form of shared or collective responsibility. Being deemed an accomplice is not as serious as being deemed a co-conspirator or co-principal, for instance.
coming his way. Perhaps a flash flood was heading for Ross’s homestead and Tally prevented a telegram from reaching Ross that would have given him time to get away. In this imagined version, Tally performs a similar action to facilitate Ross’s death at the hands of nature instead of a homicidal collective.

Consistency would seem to demand that if Tally makes himself a part of the collective action by facilitating the Skeltons’ murder in the real case, we would have to say that Tally makes himself part of a collective action with the flood in the other, and this would be odd to say the least. We do not speak of humans cooperating with natural occurrences as collective endeavors. However, in the two cases, Tally’s relationship to the Skeltons and his relationship to the flood are the same. The flood does not coordinate with Tally and neither do the Skeltons.

The problem is that without communication, recognition and coordination, there does not seem to be any difference in helping along a collection of humans and helping along a natural disaster. Both actions are a sort of risk-raising behavior that are meant to facilitate or expedite a dangerous proceeding that is already underway (Moore, 2007, p. 435). But if collective action just amounts to an aggregate of individuals who engage in risk-raising behavior by acting together, we have no need for a theory of distributing collective responsibility. Risk-raising behavior (as well as negligence or recklessness) is assessable as an individual fault, and it can be done so equally in cases of individual and collective wrongdoing. So if there is anything interesting about the idea of a group or collective intentionally performing an action together, we should maintain a distinction between individuals who engage in risk-raising behavior and individuals who intentionally cooperate in a common plan. A membership stipulation helps
maintain a distinction between genuine collective action and those (like Tally) who intentionally increase risks to others by facilitating or aiding a collective action.

The second reason I wish to maintain a membership stipulation is because it links intentional contribution to a specific domain of authorization. The problem with basing authorization on intentional contribution alone is that intentional contribution alone has no clear relationship to group authorization. The fact that an individual makes a contribution to a collective endeavor expresses very little about her willingness to have any other specific individual make a contribution toward the same end. For example, let us imagine that Bill hears on social media that there is a plan among graffiti artists to cover the local supermarket with graffiti. On his own time, Bill goes down to the supermarket, spray paints his name on the side of the building and goes home. He never communicates with any of the graffiti artists. It seems a stretch in this case to say that Bill has authorized anyone else to spray paint the supermarket on his behalf. Even if he wished or hoped that others would do so, he has not expressed his willingness in any relevant context. Nor does Bill have the intention to do so. If one tries to argue that Bill did authorize others to act on his behalf, we would have to answer the question: who did Bill authorize, all the graffiti artists, or merely some of them? This question will not admit of an easy answer if Bill is not in some way a part of a common plan of coordination with others. For this reason, intentional group membership serves to demarcate a domain of authorization. It outlines who one does and does not authorize by participating in a collective venture.

The second requirement for tacit authorization is an accurate belief stipulation. Successful tacit authorization requires not only that one join a group or common plan intentionally, it requires that one have true or accurate beliefs about the nature and relevant
specifics of the group or common plan. If one is ill or misinformed about the collective venture in which one participates, it would be implausible to assume that one tacitly authorizes one’s fellow group members to act on one’s behalf. Mistaken beliefs about an agreement are often sufficient to invalidate consent; they are therefore often sufficient to invalidate authorization as well.

Consider the following: Jared signs up for an environmental organization with the intent of taking part in an organized protest of deforestation. When he arrives at the central office, he is handed a sign and some equipment. Jared proceeds to chain himself to some old redwoods in a forest in California, to prevent loggers from cutting them down. What Jared does not know is that his act of chaining himself to a tree is a diversion for his other group members. While the foreman of the logging operation is busy dealing with Jared’s protest, the other members set fire to the foreman’s trailer and car. Jared is shocked and embarrassed. He thought this was a peaceful protest, not an act of ecological terrorism.

While it may make sense to speak of Jared as a member of the common plan, he was not an informed member of the common plan. Jared joined the organization with the intention of engaging in an activity under the description of “peaceful protest.” He did not understand his action as falling under the description of “diversion for sabotage” (Anscombe, 1957, §23). Had he known about the true purposes of his fellow group members, he would have never joined the group and chained himself to a tree. While Jared may have participated in the sabotage, he did so unknowingly. Jared therefore does not tacitly authorize his fellow group members to burn the trailer down.
It is possible that an individual could intentionally contribute to a group action while having accurate beliefs about only some of a group’s goals and activities. It is not easy to specify in advance which accurate beliefs will be relevant to group membership. The relevant facts about which one must have accurate beliefs will differ from case to case. Some individuals with a mix of accurate and inaccurate beliefs will turn out to be properly authorizing group members, while some will not. There could also be borderline cases. Let us imagine that Jared knows that the organization is an ecological terrorist organization. He chains himself to a tree, thinking that the diversion will be used by his fellow group members to burn the foreman’s trailer down. However, instead of burning the trailer down, his fellow group members detonate a bomb at the logging company’s headquarters. It is significantly harder to determine Jared’s role in the conspiracy. While he intentionally joined to be a diversion to help along wrongdoing, he did not intend the specific form of wrongdoing that occurred, nor did he have accurate beliefs about his fellow group members’ intentions.

There is typically a standard level of accurate belief that we expect of rational individuals who join organizations. A soldier may fight for a military organization that she believes to be noble and just, despite the fact that there is sufficient evidence available to the soldier to suggest that the organization is unjust and corrupt. The soldier simply refuses to attend to any of the evidence. In this case, the soldier would be excused from responsibility for the military organization’s collective endeavors on account of her false beliefs. However, she would be guilty of a different moral failing. The soldier, because she refuses to attend to the evidence against her organization, would be engaging in a form of moral negligence (May, 1992, pp. 46-52). As in law, moral responsibility for negligence increases the more “gross” the negligence.
The issue of collective responsibility becomes trickier when considering how to judge members of the common plan who are not negligently ignorant about the true nature of their group, but who remain willfully ignorant in the face of incontrovertible evidence. Imagine that the soldier does not merely ignore or omit to seek out information about her military group, but actively disavows testimony and reports about its true nature. Every time a friend or family member sends her a well sourced news article about her organization, she dismisses it as “fake news.” Upon seeing the devastating effects of her group’s activities, she comes up with alternate theories to explain away her group’s culpability. She rejects the growing concerns of other members of the group without having any good reason to do so. It is worth puzzling over whether such a willfully blind soldier fulfills the accurate belief stipulation for tacit authorization.

Willful ignorance, presumably, is a sort of epistemic vice. But the extent to which it is a moral vice will depend on the specifics of the beliefs and the social context in which they occur. Imagine that our soldier engages in mild willful ignorance. She has not done extensive research on her military organization, so she is ignorant of the illegal actions the organization plans to carry out. Were the soldier to do some investigative digging, she would uncover the truth about her organization. It simply never occurs to her to dig too deeply into what the organization does. One day, an old friend sends her a link to an article online purporting to uncover the scandalous nature of the organization. The soldier begins to skim the article. A few paragraphs in, the soldier’s tribal sentiments are engaged and she balks at the article’s content without checking closely whether or not it is true. She closes the article, assumes it is a disingenuous hit piece and never reads the article or talks to her old friend again.
In the preceding example, it is still possible to view the soldier as engaging in a form of moral negligence. For the most part, her ignorance is due to her omission to seek out the truth. She has dismissively waved off one important piece of evidence about her group, but given our best social psychology on confirmation bias and tribal sentiment, it is not too shocking that the soldier would wave off a single news clipping sent to her from an old friend. We can perhaps still view the soldier as failing to do her due diligence in researching her organization. But if the soldier remains ignorant of the organization’s plans, then her fault lies in her lack of care, not in her active coordination or cooperation. It is still possible that, were the soldier to be confronted with convincing evidence of her organization’s misdeeds, she might disavow the group and withdraw her membership from the organization.

But we could also imagine much more extreme cases of willful ignorance. What do we say about someone who ignores not a single press clipping, but every article and clipping sent to her, no matter how well sourced? We could also imagine that our soldier rationalizes away every bit of evidence that would condemn her group. When arriving upon the remnants of a massacred host of civilians, perhaps our soldier tells herself that the civilians deserved their massacre, as they must have been secret terrorists, hiding grenades and weapons under their civilian clothing. The soldier simply shrugs off every damning photograph, testimonial and official report.

I can think of two responses to such forms of willful ignorance. First, it is possible that the soldier is simply delusional. If the soldier has sufficiently lost connection with reality, it is not clear that she would be mentally sound enough to be a justifiable target of moral responsibility at all, let alone collective responsibility. We typically do not throw the severely mentally impaired into jail for their crimes, and neither do we necessarily blame people with
severe impairments for their misconduct. If our soldier is brainwashed or delusional, then she is excused from bearing collective responsibility on account of her condition. She may still be at fault for allowing herself to become delusional, but that is again a form of moral negligence or risk-raising, and not necessarily a case of collective responsibility.

Alternatively, our soldier could simply be disingenuous. Perhaps she knows deep down that her organization is corrupt and violent. She may even privately be perfectly aware of her organization’s crimes and approve of them. However, publicly, she makes sure to play the fool and pretend that she has no idea what the organization’s critics are talking about. In such a case, there is no real moral dilemma. Despite the soldier’s public behavior, she clearly fulfills the belief criterion. That she sometimes makes a show in public of not believing does not invalidate her tacit authorization. She is not delusional and, despite her public utterances, she carries on as if she understands perfectly well what her organization is about.

Willful ignorance presents some difficulties because it can manifest two epistemic conditions. Willful ignorance could be an expression of epistemic malfunction. Perhaps the person can no longer effectively form rational beliefs. In bad enough cases of epistemic malfunction, it may no longer be appropriate to speak of moral responsibility for one’s beliefs or actions based on those beliefs. Conversely, willful ignorance could be a manifestation of epistemic dishonesty. Perhaps one simply lies or deceives others about one’s beliefs in public. Epistemic dishonesty is a vice, and it does not excuse one from collective responsibility for authorizing the wrongdoing of others, assuming one meets the other relevant criteria.

I generally speak of having accurate beliefs instead of knowledge to avoid wandering into controversial discussions in epistemology. There has been a great deal of literature on the
question of what turns true belief into knowledge. In a famous paper, Edmund Gettier (1963) argued that true beliefs obtained by luck could not be knowledge. In his own example, a man forms a justified true belief from inferences based on a premise that turns out to be false. Gettier argues that the man’s justified true belief ought not count as knowledge. Others have suggested that only those true beliefs acquired in a reliable (Goldman, 1979), virtuous (Sosa, 2007, ch. 2) or sensitive (Nozick, 1981) way count as knowledge. The word “knowledge” therefore comes with philosophical baggage, and I wish to sidestep most of these issues. I will leave it as an open question to what extent one’s accurate beliefs about group membership must be formed reliably, virtuously or non-deviously.

One possible objection to the accurate belief stipulation is that it may be superfluous or unnecessary as a condition for distributing collective responsibility. For instance, Kutz (2000b, pp. 19-20) warns against requiring too steep of a belief condition for collective action. He argues that there are cases of spontaneous collective action where participants have weak to no positive beliefs about each other’s intentions (p. 20). A mob riot is a good example of such a phenomenon. A collection of separate individuals may get out their pitchforks, storm the gates of city hall and destroy the mayor’s mansion. They act together, but each member has weak or no positive beliefs about the intentions of the other members of the mob. In their fervor, they are too busy to think about anything except the impending destruction. It seems here that we have an example of collective action (and therefore targets of collective responsibility) that does not require accurate beliefs about a common plan.

Examples of mob action are by far the most difficult cases for a theory of tacit authorization. There seems to be little to no communication or recognition among mob members.
There rarely seems to be any common plan of coordination at all. It may therefore not make sense to say that mob members tacitly authorize each other to bring about a shared end. I think this is true. My response is to argue that ascribing moral responsibility for a mob action may not require a theory of distributing collective responsibility. Emerging theories in social psychology propose that participating in a mob is more akin to participating in a religious rite or dance party than a business or game of chess (Haidt, 2012, pp 268-269). One “loses” oneself in the mob, to speak metaphorically. Therefore, in many cases it makes more sense to argue that those who do wrong by certain forms of spontaneous participation are engaging in risk-raising behavior rather than participation in a collective venture.

Members of a mob might be “open” to altering their behavior based on what the other mob members do, so they do exemplify a basic form of coordination (Kutz, 2000b, p. 6). But if spontaneous group actions are by their nature in the moment and unplanned, participants cannot have strong beliefs or expectations about how (specifically) fellow mob members will act, beyond what they can predict from general beliefs about human psychology. I think it is therefore plausible to stipulate that helping a mob along is akin to helping a natural disaster along (Rescher, 1998, p. 54). For this reason, a theory of distributing collective responsibility is not always appropriate to assign fault for mob participation. Individual members of a wrongful mob action who are a party to a common plan will be subject to distributive collective responsibility, while those who are not a party to the common plan will be held to blame for risk-raising behavior. Some may find this distinction unintuitive, but there are additional normative and theoretical costs to a distributive theory of collective responsibility founded on mere intentional participation, as we shall see in chapter two.
The final requirement for the distribution of collective responsibility is that one make a substantial contribution to the relevant collective action. Marginal participation is not sufficient to demonstrate one’s willingness to others. Those who participate in a collective action marginally therefore do not tacitly authorize the other group members to act on their behalf. Once one’s contribution reaches a certain threshold of substantiality, it becomes plausible to think that one does authorize one’s fellow group members to act on one’s behalf.

The idea of a substantial contribution is, admittedly, somewhat vague. I borrow the expression from law, where it serves to pick out the antecedents that are salient contributions to the causation of some state of affairs (Hart and Honoré, pp. 123-4, 293-5). While the notion of a substantial contribution may create line-drawing issues, there is no simpler concept that can replace it entirely.\(^\text{12}\)

Some may suggest that the notion of a substantial contribution could be cached out in purely counterfactual terms. Perhaps a substantial contribution is just a contribution that stands in a relationship of counterfactual dependence to the outcome resulting from some collective action. In other words, a contribution is substantial if it is the case that had the contribution not occurred, the outcome would not have occurred either. But one can come up with counterexamples to such an interpretation. There are cases of overdetermination where individuals contribute substantially to a collective action that does not counterfactually depend on them. Consider the case of a CEO who makes corporate decisions to engage in ecologically devastating practices. Let us also assume that, if the CEO had been fired or incapacitated, some

---

\(^{12}\) For a brief discussion of how the notion of a “relevant” contribution figures in determinations of criminal liability, see Hörnle (2007, p. 145).
other executive would have made the same decision. The firm’s ecological malpractice does not counterfactually depend on the CEO, yet I think it is the case that by ordering others to enact the harmful policies, the CEO makes a substantial contribution nevertheless.

Counterfactual considerations do play a role in how we judge contributions to be substantial or inconsequential. When a specific outcome depends counterfactually upon the contribution of a single individual, we do judge that individual to be an important factor in the outcome. But in cases of overdetermination, this is not always the case. Counterfactual dependence is frequently a component of substantiality, but it is not the only component.

Gerstenberg and Lagnado (2010; Lagnado and Gerstenberg, 2015) argue that in addition to counterfactual dependence, people generally take into consideration two other factors when determining moral responsibility; in collective contexts. First, they argue that people consider not only counterfactual dependence in actuality (whether the outcome actually depended on the individual’s contribution) but counterfactual dependence in a variety of similar hypothetical situations as well (Lagnado and Gerstenberg, 2015, p. 218). We want to know if the counterfactual dependence is a fluke, or if it would also exist in a variety of similar hypothetical configurations of the world (“nearby possible worlds,” to use the language of contemporary Anglo-American metaphysics; see Lewis, 1986). The second determination is one of expectation. We tend to hold individuals responsible if it is reasonable to expect that their contribution would produce or alter a certain outcome (p. 220).

Substantiality of contribution therefore seems to hinge on a few factors. First, it depends on the scope of the individual’s causal contribution. If the agent freely and intentionally exercises a proportionately large magnitude of causal power given the context of the group
action, then the agent’s contribution is prima facie substantial. Second, substantiality depends on hypothetical facts about the agent’s relationship to the collective action. If an agent’s contribution is such that the collective action succeeds far more frequently in similar hypothetical situations in which the agent contributes, then the agent’s contribution is prima facie substantial. Finally, if an agent has authority or power such that other members of the group have a disposition to obey her and carry out her commands, her contributions to the group action will be prima facie substantial. These factors are merely prima facie because they can be necessary and sufficient or necessary but not sufficient for determining the threshold of a substantial contribution.

From these three factors, we can derive three measures for whether a contribution is substantial. The magnitude or proportion of the agent’s causal contribution determines how efficacious the agent was in the collective action. Facts about how the collective action depend on the agent’s contribution across variety of similar, hypothetical examples determine how integral the agent’s contribution was. Finally, facts about how the agent’s contribution relates to the dispositions of other group members determine whether or not the agent’s contribution was influential. It seems to me that some substantial acts will meet one criteria more patently than the other two, while others will obviously meet all three. For this reason, it is hard to give a reductive analysis to the concept of a substantial contribution (see Feinberg, 1962, p. 342).

It is often difficult to determine if a contribution is substantial, and the previous criteria help to explain why this might be the case. Consider the following example. Two men want to

---

13 Theorists who balk at talk of causal power are free to cache this stipulation out in counterfactuals, but doing so will create problems with cases of overdetermination, where it seems that the outcome of a collective action depends on none of its members individually; see Petersson, 2013.
rob a bank together. One robber is the enforcer who will hold up the bank tellers, while the other is a safecracker. The robbers run into a problem: the safe at the bank is too advanced for the safecracker. The safecracker thinks he will be unable to unlock the final locking mechanism. Because of this, the robbers call a lock-picking specialist living in another country. This lock-picking specialist is one of three people in the world who can unlock the final mechanism of the safe in question. The lock-picking specialist agrees (for a fee) to help the safecracker over the phone when he has gotten to the final locking mechanism. The two robbers hold up the bank, and the safecracker calls the lock-picking specialist when he has gotten to the final mechanism. The specialist then walks the safecracker through the final procedure on the phone.

In this example, the lock-picking specialist’s contribution is integral to the collective action. The robbery could not have gone on without him, and this is true even in related hypothetical examples, as the specialist is one of the only individuals who could provide the relevant information. However, the specialist’s contribution is not efficacious, as the other two robbers do most of the work. The specialist’s contribution is also not one made from a position of power or authority. It is therefore hard to say whether the lock-picking specialist makes a substantial contribution to the collective endeavor or not. I think reasonable people will probably disagree as to whether this is the case. Such a line-drawing problem makes sense, as the specialist’s contribution clearly meets one criteria for substantiality, but not the others.

One can imagine cases in which an individual is efficacious but not integral or influential. If a corporation instructs its low-level employees to engage in environmentally unsound activities, then the employees will be the proximate cause of the harm. Their contributions to the resulting harm could be a major causal factor. But the low level employees lack any real power.
in the group; they could do as they are told or resign. In addition, the low-level employees are not integral, for most of them are easily replaced. Do the causally efficacious low-level employees make a substantial contribution to the ecological harm? Again, I think it is hard to say and I expect that reasonable people will disagree.

To deal with problem cases in determinations of substantiality, we can use something akin to Feinberg’s (1984, pp. 150-159) solution to line-drawing concerns. We first outline three categories: (1) contributions that clearly meet all three criteria, (2) contributions that clearly meet none of the criteria and (3) those that meet only one criterion or meet all three criteria weakly. We can then sort contributions into those that are patently substantial and patently insubstantial. Finally, we decide whether or not gray area cases ought to be assigned to categories (1) or (2). The conservative approach would be to stipulate that all cases falling under (3) are insubstantial contributions. The aggressive approach would be to stipulate that all cases falling under (3) are substantial contributions like those in group (1). I tend to favor the conservative approach here, but it is not necessary to decide between the two at this point.

To summarize my argument so far, I have proposed that tacit authorization of one’s fellow group or common plan members requires membership in the group or common plan, accurate beliefs and a substantial contribution. Once these factors are met, an individual tacitly authorizes her fellow group members to carry out the relevant tasks of the common plan on her behalf. Intentionally joining a group with accurate beliefs about the common plan and making a substantial contribution toward that common plan are sufficient to signal one’s willingness to others to have them act on one’s behalf. These criteria would make it reasonable for any fellow group or common plan member to believe that one has authorized him or her to act on one’s
be half. They function to make an agreement among members of the common plan that need not be formally or explicitly expressed. Because this agreement is in place, it is therefore right to morally credit or discredit individuals for the contributions that they authorize others to carry out.

For the sake of brevity, we can formalize the tacit authorization component of a theory of distributing collective responsibility, as follows:

[TA] If a person (1) intentionally and freely becomes a member of a group or common plan, (2) possesses sufficient accurate beliefs about the group or common plan and (3) freely and intentionally makes a substantial contribution to the common plan, then that person tacitly authorizes the other members of the group or common plan to act on the person’s behalf to realize the shared aim of the common plan. If a person tacitly authorizes the other members of a group or common plan to act on the person’s behalf, then that person may be justifiably held to blame for the other members’ (relevant) actions and the reasonably foreseeable consequences of those actions.

Some may worry that TA permits too much moral luck. For those who believe moral responsibility must be (mostly) luck-free, we can also stipulate a reduced luck tacit authorization component by replacing actions with intentions to act and removing fault for the consequences of actions:

[RLTA] If a person (1) has an intention to become a member of a group or common plan, (2) possesses sufficient accurate beliefs about the group or common plan and (3) has an intention to make a substantial contribution to the common plan, then that person tacitly authorizes the other members of the group or common plan to act on the person’s behalf.
to realize the shared aim of the common plan. If a person tacitly authorizes the other members of a group to act on the person’s behalf, then that person may be justifiably held to blame for the other members’ (relevant) actions.

Because the criteria for RLTA are intentions and knowledge states in the mind of the agent, we reduce the probability that external factors will have a large effect on an agent’s moral responsibility.

My preference is for TA over RLTA, so when I speak of an authorization theory of collective responsibility throughout this dissertation, I will have TA in mind. However, nothing crucial hangs on choosing TA over RLTA. Those who dislike moral luck or are drawn to a mentalist theory of authorization are free to substitute RLTA for TA in any of the arguments that follow.

**The Moral Separateness of Agents and Proportionality**

Anyone familiar with the history of ethical and political philosophy may interject at this point and question the necessity of appealing to tacit authorization. Tacit consent is currently unpopular among political theorists, as many view Locke’s attempt to ground political obligation on the tacit consent of the citizenry as a failure (Simmons, 1976; 1979). Some may worry that tacit consent or authorization is an ad hoc device meant to arbitrarily bridge some normative gap in a deficient moral or political theory. There are some theoretical costs in introducing tacit authorization as a component of a moral theory, so it will first be necessary to demonstrate why tacit authorization should play any role in a theory of distributing collective responsibility.

I suggest that tacit authorization plays a normative role in an authorization theory of distributing collective responsibility. One compelling complaint aimed at theories of collective
responsibility is that they are intrinsically unfair or unjust (H.D. Lewis, 1948; Narveson, 2002, p. 190; Velasquez, 1983; 2003). Moral theorists in different traditions have argued for the centrality of fairness to morality (Aristotle, 1999, V.3-5; Gensler, 2013; Hare, 1972, p. 171; Kant, 2012, 4:393-405; Rawls, 1971, pp. 11-7; Haidt, 2012, pp. 158-161). If this is so, then there is something dangerously incoherent about positing a moral theory that is intrinsically unfair or unjust. I will argue that a theory of distributing collective responsibility based on authorization is capable of meeting high standards of fairness. An authorization theory can therefore address skeptical concerns about the tenability of distributing collective responsibility. Other prominent distributive theories of collective responsibility will not prove so resistant to the worries of the skeptic.

To demonstrate that a theory of distributing collective responsibility based on authorization meets high standards of fairness, I posit two criteria of fairness to normatively assess any distributive theory of collective moral responsibility. The first criterion of fairness I call, with a nod to Rawls, the moral separateness of agents. The second criterion is one of moral proportion.

Rawls proposes a notion now known as the separateness or distinctness of persons in A Theory of Justice as an objection to utilitarianism as a political theory (1971, pp. 26-7). A utilitarian theory of justice would dictate that a legislator should weigh the costs and benefits of any policy and institute those policies that would bring about the greatest net ratio of good to bad consequences. The consequences in question would be some aspect of the relevant population’s

---

14 Utilitarians who stipulate that everyone’s interests ought to be weighted equally also bake a form of impartiality and fairness into the moral cake; see Mill (2009, p. 62).
welfare, whether their psychological wellbeing or the satisfaction of their interests. To calculate the welfare impact of any policy measure, the legislator would sum up the positive and negative consequences of the policy measure on a population’s welfare, weighting every inhabitant’s welfare equally in the moral calculus.

Rawls’s worry about utilitarianism is that if each person’s interests are weighted equally, it is conceivable that there will be policies that would require a violation of the liberty of a few individuals to produce tremendously good consequences for the community at large (p. 26). In cases such as this, utilitarianism dictates that not only are such policies permissible, they are obligatory. Rawls suggests that such policy options suffer from a basic problem: they are unfair to the minority whose interests must be sacrificed for the greater good (and therefore not preferable for a group of legislators thinking about policy options behind a “veil of ignorance”; Rawls, 1971, pp. 136-142). He argues that (involuntarily) sacrificing the basic liberty interests of an individual for a greater social good “does not take seriously the distinction between persons” (p. 27).

Theorists of collective responsibility regularly invoke Rawls’s concept of the separateness of persons (Smiley, 2010; Radzik, 2001, p. 456). They claim that certain conceptions of collective responsibility ignore or efface an individual’s moral agency, because those theories hold individuals responsible for actions that they did not intentionally perform. Therefore, the argument goes, distributing collective responsibility violates the separateness or distinctness of persons.

It is not immediately clear that Rawls’s criticism of utilitarianism easily translates to problems of collective responsibility. Rawls takes issue with utilitarianism because (he argues) it
considers the basic liberty or welfare interests of each individual person to be fungible. But something slightly different is going on when we distribute collective responsibility. Theories of distributive collective responsibility do not treat the basic liberty or welfare interests of individuals as fungible goods; rather, they hold some individuals to blame for the actions of others. The issue under consideration is not basic liberty interests or welfare, but the fair ascription of moral merit and demerit.

Although Rawls develops the idea of the separateness of persons in a different context, it is possible to fashion a similar principle with respect to distributing collective responsibility. Theories of collective responsibility that, lacking a compelling normative rationale, hold some individuals to blame for the free and intentional actions of others violate what I will call the moral separateness of agents. The moral separateness of agents dictates that each individual has a right to be considered on his or her own merits. If a person acts well, she ought to be credited, and if she acts poorly, she ought to be discredited or faulted. Either way, it is prima facie unfair to blame or praise her for the free intentional actions of others. To ascribe fault to a person’s moral record for actions that she did not author disconnects her moral standing from her agency, and this is paradigmatically unfair. Such imputed ascriptions of fault (lacking any special justification) violate the separateness of agents.

The second principle of fairness I propose is proportionality. The basic idea behind the notion of proportionality is that the level of fault for an action should approximately match the gravity of one’s personal transgression (Ohlin, 2011, p. 750). Similarly, the credit ascribed to a good performance should match the nature of the praiseworthy act. For instance, making a mildly irritating and insensitive comment is a minor moral fault deserving of moderate discredit,
while a murder or rape is a grave moral fault deserving of severe moral discredit. One could say the same for moral credit as well: being polite by smiling at a neighbor deserves a lesser magnitude of moral credit, while taking an enormous risk to save a small child from drowning deserves tremendous moral credit.

A similar principle guides some notions of punishment. Some believe that in meting out sanctions, we should make sure that the punishment fits the crime (Hart, 1968, p. 160). For instance, it would be unfair to assess a life sentence to an individual who stole a box of diapers from a department store. Similarly, there is something unjust about letting a rapist off with nothing more than a warning not to engage in such behavior again. The moral principle of proportion functions similarly, but without any consideration of the social utility of punishment.

There are two components to a theory of moral proportion. The first component is that of desert. An individual has a right to be considered on her own merits or demerits, but she also has a right to a fair reckoning. An individual therefore deserves to be credited a level of praise in proportion to her good conduct and deserves to be credited a level of blame in proportion to her bad conduct. In the words of legal theorist Frédéric Mégret, individuals have a right to a “fair labeling” of their actions (Ohlin, 2011, p. 751). The second component of a theory of proportion is consistency. Consistency requires that the acceptable proportions of blame and praise for an act-type should be applied consistently to individuals and between individuals. If making a crude joke is a minor infraction for person A in a specific social environment, then making a crude joke should be a minor infraction for person B in the same social environment, other things being
equal. Similarly, if it is a major moral fault for A to murder B at t1, then it should be a major moral fault for A to murder C at t2, other things being equal.\footnote{Frequently, other things are not equal. It may not be a moral fault for an individual to crack a joke about her own ethnicity. However, if a member of another ethnicity makes the same joke, it could very well be morally blameworthy.}

To get an intuitive sense of how proportionality works, we may reflect on how children are praised or scolded for their behavior. If a parent punishes one child severely for forgetting to wash her hands before dinner while barely criticizing another child for intentionally harming and sending another child to the hospital, we would think that the parents are being unfair. This is because the parents violate norms of proportionality; not only are their reactions disproportionate to the magnitude of the children’s fault, their reactions are inconsistent between the two children.

The moral separateness of agents and proportion serve as two principles of fairness to normatively assess any theory of distributing moral responsibility\textsuperscript{f}. Any theory that violates one or both principles, whether or not it has social utility, will be shown to be prima facie unfair. Moral theorists therefore have a normative reason to prefer any theory that can satisfy the demands of both principles.

**What Fairness Requires**

I claimed earlier that the advantage of a theory of distributing collective responsibility\textsuperscript{f} based on tacit authorization is that it is capable of meeting high standards of fairness. In this final section, I begin to make good on this claim. I demonstrate that an authorization theory meets the demands of both the moral separateness of agents and proportionality. I finish by briefly situating an authorization theory of distributing collective moral responsibility\textsuperscript{f} in the larger individualist/collectivist debate on collective moral responsibility\textsuperscript{f}. 

---
Recall first the principle of the moral separateness of agents. The moral separateness of agents stipulates that every agent ought to be judged on the basis of his or her personal conduct. It is unfair to blame (or praise) an individual for actions that stand in no relationship to her personal decisions and actions. To ascribe blame or praise without taking note of an individual’s intentional actions decouples the strong relationship between responsibility and autonomy. By maintaining the connection between autonomy and responsibility, we ensure that individuals only receive the credit or discredit that they deserve. We also ensure that individuals are treated fairly in the distribution of responsibility.

A common criticism of theories of collective responsibility is that, because they hold some individuals responsible for the actions of others, they are intrinsically unfair or unjust. Consider one of the most traditional and controversial forms of distributing collective responsibility: guilt by association. Guilt by association effectively transfers the fault of some group member to all other members of the group. One becomes at fault for group wrongdoing simply by being part of the group. Guilt by association usually makes no distinction between groups that require intentional membership and those groups constituted by identity. One must answer for the transgressions of one’s corporation, just as one must answer for the transgressions of one’s race or religion.

Many now find guilt by association to be repugnant, and it is rightfully decried as being unfair. Guilt by association clearly violates the moral separateness of agents. Group members who did not partake in collective wrongdoing in any way are held to blame for the actions of others. According to guilt by association, responsibility is only sometimes a matter of intentional action. Some individuals are blamed for the conduct they intentionally perform, while others are
blamed for standing in a relationship to those who commit wrongdoing. Guilt by association even holds small children and the physically incapacitated to blame for the actions of their fellow group members. Because guilt by association unfairly ascribes responsibility to those who have personally done nothing at all, it clearly violates the moral separateness of agents.

Unlike guilt by association, a theory of collective moral responsibility founded on authorization does not violate the moral separateness of agents. The problem with guilt by association is that it distributes fault to individuals without any consideration of their wills or actions. But authorization relationships are constituted by the free and intentional actions of individuals working to further a collective end. On the theory I propose, fault only distributes to members of a collective who expressly authorize their fellow group members to act or whose contribution is sufficient for tacit authorization. Collective fault is therefore always a function of an explicit agreement or (intentional) substantial contribution. This serves to maintain the strong link between individual autonomy and responsibility. If someone is held to blame, it is because of something she has freely and intentionally authored, either through her own person or vicariously through an agreement with others.

An authorization theory of distributing collective moral responsibility also satisfies demands of proportionality. Proportionality requires that the magnitude of credit or discredit ascribed to an agent be proportionate to the magnitude of her wrongdoing. Because tacit authorization requires a substantial contribution for the distribution of fault, it avoids cases where marginal contributors are held to blame for large collective actions. Marginal contributors to collective actions do not tacitly authorize their fellow group members to act on their behalf (although they may sometimes explicitly do so). According to a theory of distributing collective
moral responsibility based on authorization, fault for collective wrongdoing distributes to those who have successfully authorized others to vicariously act on their behalf. But those who authorize their fellow group members effectively make themselves the authors of the collective action in question through a collection of authorization agreements. It is therefore not disproportionate to hold the authorizers of group actions to blame for large collective outcomes. By vicariously producing an outcome through others, it is as if the authorizers had performed the actions themselves.

If we want an example of a theory that fails to meet the demands of proportionality, we can return to guilt by association. According to guilt by association, group membership alone is sufficient to distribute collective fault. Consider a janitor who mops the floors at a large corporation engaged in ecologically destructive activities. Let us assume that the janitor knows more or less about the devastating effects of the company’s activities. He is also not coerced by social or economic factors to keep working for the corporation. The janitor continues to work for the company and he intentionally contributes to its proper functioning. Guilt by association would attribute fault for the harm caused by the entire operation to the janitor. But that seems to assign a tremendous magnitude of fault to an individual who contributes very marginally to the production of the harmful outcome. To ascribe the entirety of the harmful outcome to the janitor would in effect say that his marginal participation is equivalent to him personally causing the ecological devastation himself. This seems incorrect; if it were correct, it would become very difficult to distinguish between the blame due for minor infractions and serious transgressions, as any minor infraction could in principle merit the same magnitude of fault as a devastating act of true evil.
Conclusion

Similar to Locke’s account of tacit consent, the tacit authorization component of a theory of distributing collective responsibility bridges a normative gap. The normative gap is how we go from blaming individuals for outcomes due to their personal conduct to blaming individuals for harmful results produced by others. The theory I propose stipulates that we may hold those individuals to blame who vicariously author a wrongful outcome through authorization agreements. If an individual is not the (personal or vicarious) author of an outcome, absent any existing forward-looking obligations, we have no moral reason to hold her to blame for the outcome’s production. Authorization therefore fills out a normative story of how we get from holding individuals to blame for their own conduct to holding individuals to blame for the conduct of others.

I have offered a theory to explain the distribution of fault among group members. The astute observer will note that in doing so, I have left one of the most controversial questions in the literature on collective responsibility untouched. I have not addressed the matter of whether or not collectives themselves might be rightfully held to blame for their actions. Some may worry that I have therefore missed something crucial about the concept of distributing collective responsibility.

But an authorization theory can remain neutral on the ontological status of collectives. I have so far presented a *distributive* theory of collective responsibility in an individualist form. Group fault distributes to those members of a collective who expressly or tacitly authorize the other group or common plan members to act on their behalf. But there is no reason that one may
not also authorize the collective itself. One could stipulate that by meeting the threshold of tacit authorization, one not only authorizes one’s fellow group members to act on one’s behalf, one authorizes the collective itself to act on one’s behalf. This is not so strange an assertion, for individuals regularly authorize corporations to make decisions for them. Authorization relationships build normative connections between one individual and others. The relationship in question could be one of a single individual authorizing another, but it could as well be a single individual authorizing a collective in its entirety.

With this said, it is a virtue of an authorization theory of distributing collective moral responsibility that it is potentially elimitivist. Because tacit and express authorization forges authorization relationships between individuals, we can potentially explain collective moral responsibility without appealing to the existence of substantial collectives. If authorization tells the whole story of what is interesting about collective responsibility, then we could do away with the need for substantial collectives in determining how collective responsibility distributes, on the ground that an authorization theory of collective responsibility alone is more conceptually parsimonious than one requiring the existence of substantial collectives.

But I have not come close to demonstrating that an authorization theory of distributing collective moral responsibility is capable of eliminating any need for appeal to substantial collectives. Such a project would require a serious look at the metaphysics and semantics of collective action. My goal in this chapter has been to demonstrate that authorization solves a normative problem in discussions of how to fairly distribute collective moral responsibility. Authorization accounts for how fault distributes from one individual (or individuals) to another. It does so in a way that respects the separateness of moral agents and avoids violating principles
of proportion. It is therefore a method of distributing responsibility, that can meet rigorous standards of fairness.

I have so far argued for a specific understanding of an authorization theory of distributing collective moral responsibility, and that such an account avoids charges that it may be intrinsically unfair. However, while I have demonstrated that an authorization theory may be viable, I have not yet demonstrated that it is preferable to other distributive theories of collective responsibility. In the following chapters, I compare an authorization theory with prominent competing alternatives in the literature. In each case I will argue that an authorization account of collective moral responsibility is preferable, either because it more successfully answers charges of unfairness or because it is not subject to the same conceptual worries.
Many theorists of distributing collective responsibility$_f$ working in the analytic or Anglo-American tradition gravitate toward what I will call an intention theory of collective moral responsibility$_f$. An intention theory of individual responsibility$_i$ holds that moral responsibility$_i$ for a morally prohibited act derives from the individual’s intention to personally carry out that act. An intention theory of distributing collective responsibility$_f$ holds that the possession of a specific sort of intention or set of intentions to do wrong is necessary (and, depending on the theory, sufficient) for the distribution of collective fault to the individual. If one intends or aims at an unjustified collectively produced outcome, then one is prima facie to blame for that outcome.

One influential intention account of distributing collective responsibility$_f$ comes from Kutz (2000a). Kutz argues that individuals share responsibility for actions performed by others when they participate intentionally in a collective activity (p. 122). In his own example, Kutz argues that all the bombers who participated in the firebombing of Dresden share responsibility for the full consequences of the fiery inferno (p. 141). Each bomber instantiates what Kutz denotes a “participatory intention” to commit wrongdoing (p. 74). A participatory intention has two forms of representational content: one must correctly conceive of a collective action as

---

1 Portions of this chapter have been reworked and published as Atenasio (2018).
requiring the intentional participation of others and one must understand one’s contribution as furthering the collective end (pp. 81-82). Once these criteria are met, one bears personal responsibility not only for one’s own contribution to the collective action, but the contributions of one’s fellow participants as well.

Kutz refers to his theory of participatory intentions as a minimalist theory. But there are more complicated or sophisticated intention theories of distributing collective responsibility as well. Tuomela and Mäkelä (2016) argue that collective responsibility distributes to all members of a group who have what they call a “collective intention” to commit wrongdoing (p. 307). In an earlier work, Tuomela (2006, p. 43) argues that a collective intention obtains whenever an individual intends to participate in a collective action, believes that she will have the opportunity to make a contribution to the collective action, believes that other members of the collective action will do their part and believes that the group members are aware of each other’s aims to carry out the collective action. Once these conditions are met, the individual instantiates a collective intention and is liable to be blamed for the contributions that her fellow collective members take to achieve the common end.2

If Kutz’s minimalist intention theory of distributive collective responsibility proves viable, it is bad news for an authorization theory.3 Many philosophers (and scientists) take simplicity and conceptual parsimony to be theoretical virtues (see Quine, 1966). Without further

---

2 The reduced luck tacit authorization theory (RLTA) that I propose in chapter one, requiring only a cluster of mental states for the distribution of collective responsibility, could also be thought of as a sophisticated form of intention theory.

3 Recall that the authorization theory of collective moral responsibility I defend in chapter one requires intentional and uncoerced group or common plan membership, accurate beliefs about the group or common plan and an intentional and uncoerced substantial contribution to further the group action or common plan for the distribution of collective fault to the individual.
examination, minimalist intention theories such as Kutz’s appear to be simpler and more conceptually parsimonious than either authorization theory I propose. For the minimalist intention theorist, collective responsibility is a function of a single mental state: an intention to participate in a collective endeavor aimed at a shared end. If a single intention adequately justifies the distribution of collective responsibility, then more complex theories of distributing collective moral responsibility, such as an authorization account, will be shown to be conceptually superfluous (Kutz, 2000a, p. 226).

My goal in this chapter is twofold. First, I argue that Kutz’s minimalist intention theory of collective responsibility leads to troubling results. While one can argue that his intention theory does not violate the moral separateness of agents, it nonetheless violates norms of proportionate blame. If intentional participation entails fault for a collective outcome, then marginal participants taking part in a collective endeavor may be to blame for the catastrophic consequences of large group actions. According to Kutz’s theory, even the janitors who intentionally took part in the allied war effort bear responsibility for the consequences of the Dresden firebombing. Such extreme attributions of fault to marginal participants who play a small role is disproportionate and unfair.

Some have proposed solutions to mitigate the apparent unfairness of a minimalist intention theory of collective responsibility. Kutz himself proposes that we distinguish between principal wrongdoers and accomplices, as well as inclusive authorship and exclusive authorship (2000a, p. 146). Gadirov (2011, p. 15) suggests that only group members who have closely interlocked intentions become responsible for the contributions of their fellow group members. Sophisticated intention theories such as Tuomela and Mäkelä’s add additional mental state
requirements for the distribution of collective responsibility, thereby preventing the distribution of fault to marginal participants in the collective action.

My second claim is that these modifications and more sophisticated intention theories fail, because they do not offer a compelling rationale to justify expanding a far less controversial notion of responsibility (responsibility for one’s own actions) to include a far more controversial concept of responsibility (responsibility for the actions of others). Intention theorists typically defend this expansion on the ground that the possession of a “collective intention” is sufficient to normatively justify the distribution of collective responsibility. Or, intention theorists argue that it is necessary satisfy “our” intuitions or attitudes about collective wrongdoing (Kutz, 2000a, p. 122; Mathiesen, 2006, p. 242; Silver, 2002, pp. 291-294; Smiley, 2010, p. 198). But I argue that these strategies fail. Until intention theorists can provide a more compelling rationale for expanding the principle of individual responsibility to include responsibility for the actions of others, we ought to refuse this expansion.

Conversely, an authorization theory of collective responsibility offers a compelling rationale to justify holding some individuals responsible for the actions of others. When one person authorizes another to carry out wrongdoing on her behalf, we rightfully blame the authorizer for the actions performed by her agent. We do this because the authorizer has freely and intentionally entered into an agreement with the agent to further her interests. The authorizer thereby acquires moral responsibility for what her agent does on account of the agreement. Because an authorization relationship is morally transformative between two individuals, we have a compelling rationale to explain how one individual might be to blame for the actions of another.
This chapter proceeds as follows. First, I give a brief overview of some developments in analytic philosophy regarding participatory and collective intentions and how they function in human action. I then highlight how discussions of collective action transition to discussions of collective responsibility. Next, I articulate Christopher Kutz’s (2000a) theory of participatory intentions. I give some reasons why theorists have found his model attractive. I then offer two objections: the objection that Kutz’s theory of participatory intentions ascribes disproportionate magnitudes of blame to marginal participants and the objection that more sophisticated intention theories of distributing collective responsibility have yet to provide a compelling rationale to justify holding some individuals responsible for the actions of others. I conclude that these objections are worrisome enough that we ought to prefer an authorization theory of distributing collective moral responsibility to the prominent intention theories in the literature.

**From Individual to Collective Action**

In her famous paper on modern moral philosophy, Elizabeth Anscombe lamented the fact that philosophers had yet to provide a coherent “philosophy of psychology” (Anscombe, 1958, p. 4). More specifically, Anscombe asserted that no theorist had successfully articulated the relationship of intentions to human action (p. 5). Anscombe herself attempts to elucidate this relationship in *Intention* (1957). Anscombe argues that an action is intentional under a given description if the agent could give a specific kind of answer to the question “Why did you do that?” (1957, §5). If the answer to the “why?” question gives a reason for the agent’s purposive behavior, then the agent acted intentionally under that description. If we ask a woman why she is feeding the ducks by throwing bread to them and she answers “I like watching them eat,” then we know that her action is intentional under the description “feeding the ducks.” But the same action may not be intentional under a different description, such as “annoying the park ranger.”
If we ask the woman “why are you annoying the park ranger?” she will not respond with a reason for her action, but with “I did not know that feeding the ducks annoys the park ranger.” The woman’s action is therefore not intentional under the description “annoying the park ranger.”

Many theorists puzzled over what Anscombe meant by a “reason for action.” Donald Davidson (1963) suggests that a reason for action consists of “a belief and an attitude” (p. 688). More specifically, Davidson argues that a reason for action is a “pro-attitude” toward a state of affairs coupled with a belief that one’s behavior aims to realize that state of affairs (p. 687). The belief and pro-attitude help us to rationalize the actions of others (p. 686). To give an example, we might ask the woman throwing bread at the pond why she is throwing bread. She answers “because I want to feed the ducks.” In this case, the woman has given us her reason for throwing bread to the ducks. According to Davidson, her reason is a pro-attitude (a desire to feed the ducks) coupled with a belief that by throwing bread, she would be feeding the ducks. The belief and desire help us understand the woman’s rationale in throwing the bread.

More recently, Michael Bratman has argued that intentions are not beliefs and desires, but a specific form of planning state (1987, pp. 9-11). If an agent intends to perform some action in the present or future, then the action fits into or relates to the content of one of the agent’s plans or goals in some way. Humans structure their lives according to these plans. If an agent performs an action because it relates to her plans, then the agent acts intentionally. If the agent does something that cannot be related to any of her plans or goals, then the agent has no intention in acting (p. 119). Bratman calls this a “planning theory” of action.

Whether intentions are belief-desire pairs or planning states will not have a particularly large impact on the nature of collective action and its relationship to collective responsibility.
will follow Bratman’s usage in this chapter: I define intentions as goals, aims or planning states. We can think of these aim or planning states as part of the agent’s psychology. An intention is therefore a mental plan or aim the individual has to carry out an action or set of actions, either presently or in the future. While there are other defensible definitions of intention, this definition is consistent with all of the prominent intention theories of distributing collective moral responsibility.

Philosophers of action typically split up planning states into two categories. First, there are direct intentions or plans to personally carry out some course of actions (Tuomela, 2006, p. 37; Kutz, 2000a, p. 87). Intentions in this category are plans that can be successfully carried out by the agent alone. For instance, I might form a direct intention to mow my neighbor’s lawn. These intentions have as their content a state of affairs that can be realized through the agent’s personal efforts. Second, there are broader aim or executive intentions (Tuomela, 2006, p. 37; Kutz, 2000a, p. 87). Aim-oriented intentions may involve states of affairs that are external to the agent’s personal actions. One could, for instance, intend to save the environment. Presumably, no individual person can accomplish this alone. The content of such an intention state would therefore involve beliefs about how others will act and cooperate. The personal role of a broader aim intention to save the environment would be the direct intention to do one’s part in saving the environment by recycling, developing sources of clean energy, and so on.

So far we have only touched upon individual intentions or reasons to act. But discussions of individual intentions have broadened to include the notion of collective or group intentions to act together. Talk of collective intentions may conjure up images of a supra-human collective mind that has its own beliefs, desires and plans. But this is typically not how theorists of distributing collective responsibility understand collective intentions. Philosophers of action
define collective intentions either as individual intentions of a special type or as regular individual intentions related to other agents and intentions in a certain way. Following Tuomela and Miller (1988), we can call the collective intentions of a special type “we-intentions,” and following Kutz (2000a), we can call individual intentions related to larger goals “participatory intentions.”

The first way theorists have understood collective intentions is as a singular form of aim or goal state. Usually, an individual forms a goal or plan and carries it out herself. These sorts of goals and plans are individual or “I” intentions. However, sometimes individuals form goals or plans together. One can think of these goals or plans as “we” intentions. We-intention theorists hold that while we-intentions are planning states in the minds of individual agents, they are planning states that are typologically different than individual or I-intention planning states (Searle, 1990, pp. 404-5; Tuomela, 2006, p. 45). Therefore, depending on whether an agent plans to act alone or with others, the agent will instantiate different types of mental states.

We-intention theorists hold an irreducibility thesis when it comes to collective intentions. They argue that collective intention states cannot be reduced to or explained away as a mere collection of individual intention states (Searle, 1990, p. 404). Rather, we-intention theorists suggest that we-intentions are a *sui generis* form of mental state that are typologically distinct from individual I-intentions.

Other theorists of collective action disagree. They argue that collective intentions are individual intentions instantiated in specific relationships. In (1993), Bratman argues that a collective intention obtains when two or more individuals intend to carry out an action J, have coordinated sub-plans to J, intend to J because the other individuals in the group intend to J and have common knowledge of each other’s intentions (p. 106). For Bratman, a collective intention
is not a special type of intention, it is a normal individual intention that stands in a set of relationships to the plans of other cooperating group members.

Kutz’s (2000a) deflationary account of collective action is even simpler. Kutz argues that collective action is a function of individual intentions to participate in a collective endeavor (pp. 89-96). For Kutz, an individual becomes a party to a collective action by intending or aiming at a collectively produced outcome. When a soldier intends to participate in the expulsion of an ethnic minority from a village, the soldier has an intention to participate in the removal of every minority member from the village, even if he knows he will only play a small part in the collective action. Like Bratman, Kutz defines a collective intention as an individual intention that stands in a relationship to a collective endeavor. Unlike Bratman, Kutz requires only one relationship between an intention and a collective activity to distribute collective responsibility: the relationship of intended participation.

Kutz calls these individual intentions to participate in a collective endeavor “participatory intentions” (p. 67). Unlike we-intentions, participatory intentions are typologically indistinguishable from individual intentions; they are run-of-the-mill individual intentions that have as their content a collective activity. Theorists who follow Kutz therefore deny the irreducibility thesis. They claim that collective intentions are nothing over and above a set of type-individual intentions instantiated in a certain relationship. In other words, if one aims to participate in a cooperative activity with others, one collectively intends the outcome of the cooperative activity.

I will argue that both Christopher Kutz’s (2000a) theory of participatory intentions and Tuomela and Makela’s (2016) we-intention account of distributing collective responsibility...
prove inadequate. However, before I can make this argument, it will be helpful to take a look at how collective or participatory intentions lead to the distribution of collective responsibility.

**From Collective Intentions to Collective Responsibility**

Talk of collective action and collective intentionality inevitably leads to discussions of collective responsibility. When an individual intentionally commits wrongdoing, we generally hold her at fault or to blame for the wrongdoing. Some argue that by parity of reasoning, if an individual or group acts on a collective intention, we should hold the individual or group at fault or to blame for the collective transgression.

Some theorists of collective action take no clear position on the existence of collective responsibility (Bratman, 1993; Searle, 1990). Others admit that collective wrongdoing may entail fault for the individual, but they claim that there is no easy formula to determine how this works (Gilbert, 2006, pp. 109-13). Some theorists of collective responsibility stipulate that collective intentional action entails moral responsibility for the collective itself, but that we cannot make any easy inferences from the collective’s fault to the fault of its individual members (Cooper, 1968, pp. 262-263; French, 1984, p. 14; Copp, 2006, p. 215-216; Pettit, 2007, pp. 180-184; Shockley, 2007, p. 451). These theorists propose that collective action exists, but that it is conceptually distinct from questions of how one individual becomes at fault for the actions of a fellow group member.

The aforementioned theories will not help us understand how fault distributes to group members, so we must turn to intention theorists who offer a theoretical framework for the distribution of fault. The most prominent intention theory to date comes from Christopher Kutz’s *Complicity* (2000a). As noted earlier, Kutz proposes a “minimalist” theory of collective intentional action (p. 90). He argues that a collective intentional action is made up of participants
who individually intend a collective goal (p. 103). If A, B and C each intentionally participate in an endeavor to build a treehouse together, then A, B and C jointly build the treehouse. The individuals need not have strong beliefs or desires about how their fellow group members will act. Indeed, Kutz argues that A, B and C need not have a great deal of determinate knowledge about what their fellow group members are doing (p. 90). According to Kutz, if an individual intentionally participates in a collective activity with others, then the individual becomes an intentional author of the entire activity.

Kutz describes intentional participation as doing one’s part in a collective act (p. 81). When one intends to do one’s part in a collective act, one’s intention has two kinds of representational content (p. 81). First, one has a conception of the collective action as collective. One understands a collective endeavor, such as going for a walk together, as something that requires the cooperation and intentional participation of others (p. 82). Second, one has a conception of one’s individual action as contributing to the realization of the collective end (p. 82). Once these two characteristics are present, one instantiates a participatory intention to bring about a shared end and one makes oneself a co-author of the collective action.

According to his theory of collective action, Kutz develops a parallel theory of distributing collective responsibility. He argues that all who intentionally participate in an unjust collective action are personally responsible for the collective injustice (p. 122). More

---

4 Kutz uses the term “accountable” instead of responsible. According to Kutz, a person is accountable for a wrong if it is right to blame that person for the wrong and expect her to compensate any victims for relevant harms or damages. In chapter one, I chose to keep fault and accountability (or liability) to repair distinct. Although the two frequently overlap, sometimes they do not. One may be accountable or liable for a wrong, even if one is not at fault for it. Parents may be accountable for the bad behavior of their children, even if they are not rightfully blamed for their children’s misdeeds. However, Kutz clearly believes that if one intentionally participates in wrongdoing, one is not only accountable or liable to repair the harm, but at fault or to blame for the wrongdoing as well; see Kutz (2000a, pp. 122, 199).
specifically, the members of a collective are all individually responsible for the actions of their fellow group members if each has participated intentionally in a collective activity aimed at a shared end (p. 122). If each group member’s conception of the collective goal overlaps sufficiently, Kutz argues that “the will of each is represented in what each other does qua group member” (p. 141). Because “the will of each [is] manifested in the acts of all,” it is correct to hold every participant responsible for the goal-oriented actions of every other participant in the collective action (p. 142).

In defense of his theory, Kutz stipulates a “complicity principle” that accounts for how one person becomes responsible for the actions of another:

**The Complicity Principle:** I am accountable for what others do when I intentionally participate in the wrong they do or harm they cause. (p. 122)

The complicity principle dictates that whenever one participates in any wrongdoing, whether small group actions or larger, global actions, one becomes complicit in the wrongdoing of the other participants. Kutz defends the complicity principle on the grounds that it is necessary to satisfy “our” emotions and considered judgments (intuitions) about collective wrongdoing. When we take up a “second-person perspective” and consider what it is like to be the victim of a coordinated, collective harm, we are led to judge that some sort of complicity principle is necessary to do justice for the full extent of the collective harm (p. 123). Were participants of the collective harm simply let off the hook because they were mere participants, this would ignore the suffering of those who were harmed by the collective act (p. 123). To refuse to hold participants in wrongdoing fully accountable for their collective aims would be an affront to the dignity of victims who suffer due to collective wrongdoing.
Kutz’s intention theory of collective moral responsibility has some clear advantages. First, it is a conceptually parsimonious theory. The distribution of collective fault results from the instantiation of a mental state that stands in a single relationship to a collective activity. If one has a conception of an activity as a cooperative endeavor and a conception of one’s own actions as furthering or contributing to the endeavor, then one participates intentionally in the collective action and bears responsibility for the actions one’s fellow group members take to promote the collective end. Other individualist theories of distributing collective responsibility require extra stipulations (see Tuomela and Mäkelä, 2016).

Second, Kutz’s theory is capable of accounting for how collective action works in a variety of contexts. Whereas more stringent theories of collective action struggle to explain how mob action and spontaneous projects count as genuine instances of collective action, Kutz’s theory easily accounts for such examples. Even if members of a mob do not have strong beliefs about how other members of the mob will act, they all participate intentionally in a collective endeavor. Similarly, two people may spontaneously decide to bring a piece of electronics inside when it starts to rain. Each person intends to participate in the shared aim of saving the electronic equipment, even if both members have not worked out a clear plan as to how they will do this. Given Kutz’s theory of participatory intentions, both the members of the mob and the individuals saving the electronic equipment count as taking part in a collective action. Theorists of collective action who require stringent criteria for the instantiation of a collective intention may not be able to account for such spontaneous or mob action.

Finally, Kutz’s theory of participatory intentions generates a theory of distributing moral responsibility that does not sever the connection between an individual’s will and the actions she performs. Unlike a theory of guilt by association, Kutz’s theory of participatory intentions
stipulates that any individual responsible for collective wrongdoing has in some way willed the wrongdoing through her intentional participation. The collective action is part of the content of her intention states, so we do not violate the separateness of agents by blaming her for the cooperative contributions of others.

While Kutz’s theory of participatory intentions has its virtues, I will argue that we still have reason to prefer an authorization account of distributing collective moral responsibility. First, without appeal to extra (ad hoc) qualifications, Kutz’s theory of participatory intentions assigns disproportionate magnitudes of blame to marginal participants. However, more sophisticated intention theories fare no better, as proponents of sophisticated intention theories have yet to offer a compelling, non-question begging rationale for assigning responsibility to some for the free and intentional actions of others. I proceed to articulate these objections in greater detail.

**Participatory Intentions and Proportionate Blame**

Recall that in chapter one I posited two standards of fairness: the moral separateness of agents and proportionality. The moral separateness of agents stipulates that it is unfair to hold individuals to blame for actions that bear no relevant connection to what they intentionally do. Intention theories do not violate the separateness of agents. Intention theories distribute blame on account of an intention to engage in wrongdoing with others. An agent is only held to blame if she aims at a collectively produced outcome in some way. Since the collectively produced wrong is part of the content of the individual’s intentions, it does not violate the moral separateness of agents to hold the individual to blame for the resulting injustice. In other words, by holding an individual who intends or aims at a collectively produced wrong to blame for the collective
action, we do not efface the connection between what an individual intends or aims at and what we hold the individual responsible for.

But there is another principle of fairness: proportionality. Proportionality demands that we assess a reasonable magnitude of credit or discredit to an agent depending on what she does. Here, minimalist intention theories like Kutz’s run into some trouble. With too few restrictions on who counts as a participating group member, minimalist intention theories fault marginal contributors for the full extent of large, global wrongs. It is disproportionate and therefore unfair to hold a marginal contributor to blame for a large, collectively produced outcome. This provides a normative reason to reject any theory of collective responsibility that holds marginal participants to blame for large, wrongful collective actions or outcomes.

Consider an example employed by Christopher Kutz to argue in favor of his theory of participatory intentions: the firebombing of Dresden (2000, pp. 115-122). The firebombing of Dresden was a truly collective action, for the overall destruction of human life was the result of a raging inferno that could only have been created as a result of a coordinated bombing effort. No individual humans could have created such destruction on their own. It is difficult to assign individual responsibility to the pilots and generals who organized and carried out the attack, as the single deadly fire was a combined effort of many people.

Kutz argues that any individuals who participate in an atrocity (such as the bombings of Dresden or Tokyo) with the relevant participatory intention bear responsibility for the atrocity (p. 141). Kutz finds this intuitive when talking of a group of bombers who coordinate to destroy a target or set of targets. He writes that “what the bombers do together is the object of the will of any given bomber, and so what the bombers do together is a potential object of inclusive individual accountability” (p. 141). Kutz argues that, because all the bombers individually have a
conception of their participation as fitting into the larger project of the bombing of the city, all can rightly be faulted for the ensuing destruction.

While such an ascription of responsibility may seem intuitive to some in the case of the bomber pilots or generals, there are many more individuals who contributed to the war effort. Mess hall attendants and cooks contributed intentionally to the bombing campaign by helping to feed the bomber pilots. Workers in ammunitions factories intentionally produced weapons they knew would be used to kill the enemy. Intelligence agents gathered information to facilitate the bombing. Dispatchers and radio operators helped to coordinate the attack.

At the level of mere participation, we can include many marginal participants as part of the bombing attacks on Dresden (or Tokyo). Given only a minimalist intention theory of distributing collective responsibility, why should these marginal participants not bear moral responsibility for the outcome of the attacks? Without further qualifications to Kutz’s theory of participatory intentions, it looks like these marginal participants are equally liable to collective responsibility as those who made a larger contribution. According to his theory, marginal participants who participated in the bombing attacks with a conception of their participation as contributing to the collective effort deserve some form of personal responsibility for the ensuing destruction. Without a clear reason to exclude some participants rather than others, we must admit that radio operators, cooks and factory workers are equally as liable for the bombing efforts as the officials who ordered the operation, as long as they intentionally participate in the war effort with the understanding that their contributions will help facilitate the operation.5

---

5 The obvious solution would be to base magnitude of responsibility on causal contribution, but Kutz rejects this line of thinking. Kutz argues that in overdetermined collective actions, such as the bombing of Dresden, we cannot rightly assign causal responsibility to anyone, as no individual’s personal contribution makes a difference to the outcome; see Kutz (2007). For a response see Petersson (2013).
But to say that a worker in a munitions factory may be equally liable to the distribution of collective responsibility\textsubscript{c} as a general who plans and orders the attack violates at least two norms of proportion. It first of all assigns too great a magnitude of fault to the factory worker, whose contribution is extremely remote from the consequences of the collective bombing. It also unfairly assigns equal magnitudes of fault to the general and the factory worker, even though the general’s participation is much more efficacious, integral and influential to the collective action.\textsuperscript{6} Both of these results should worry the theorist of collective responsibility\textsubscript{c} who is at all concerned to produce a theory of distributing collective responsibility\textsubscript{c} that aims at treating all agents fairly.

Kutz himself seems to realize this, as he amends his own theory to address worries of disproportionate blame. Kutz introduces a qualification to his intention theory of distributing collective responsibility\textsubscript{c} to distinguish those who do wrong by “direct action” from those who do wrong by “complicit participation” (p. 146). According to Kutz, those who do wrong by direct action merit “exclusive” responsibility\textsubscript{c}, while those who do wrong by complicit participation merit “inclusive” responsibility\textsubscript{c} (p. 146). Kutz believes that both types of responsibility\textsubscript{c} are legitimate forms of moral responsibility. However, Kutz argues that individuals may be due different moral or punitive responses if their moral responsibility\textsubscript{c} is inclusive as opposed to exclusive, although he does not elaborate on how this is supposed to work (p. 165). Kutz rather argues that there is no clear guiding principle to tell us how responses ought to differ or what they ought to be (p. 148). He believes specific determinations must be made on a case by case basis.

\textsuperscript{6} See chapter one for a definition of these terms.
Gadirov (2011) also argues that a minimalist intention theory needs to be supplemented to avoid assigning disproportionate blame. He believes that in addition to intentional participation, group members must demonstrate “meshed goals or a web of interlocked intentions” to merit fault for collective wrongdoing (p. 15). The idea of meshing goals is one he borrows from Bratman (1993). Intentions or goals mesh when members of a collective action more or less agree upon the technical specifics of how the collective action ought to be carried out. For instance, if agents A and B intend to collectively paint a house blue, then A and B’s plans and sub-plans will mesh only if both intend to paint their portion of the house blue (Bratman, 1993, p. 106). If A intends to paint her portion red, then their intentions will not mesh. Gadirov argues that mere intentional participation is not enough to merit responsibility for a collective wrong; rather, only those participating agents with sufficiently determinate and coordinated plans will reach the threshold of acquiring responsibility for the contributions of their fellow group members.

Such modifications would prevent minimalist intention theories from assigning blame too liberally. If we stipulate that marginal participants in a collective wrong merit a lesser form of responsibility, we can explain why integral group actors deserve greater blame for collective wrongs than marginal participants. Integral group actors deserve exclusive responsibility, which presumably is a more serious or blameworthy form of moral responsibility than the inclusive responsibility due to marginal participants. Alternatively, we could stipulate that marginal participants in collective wrongdoing do not deserve personal responsibility for the full outcome of the cooperative endeavor because their goals or plans do not coordinate or mesh determinately. This would also explain how integral group actors are due more blame than marginal participants.
While such qualifications may help minimalist intention theories of distributing collective responsibility avoid charges of disproportionate blame, they defeat one of the primary reasons for favoring a minimalist intention theory. A minimalist intention theory’s primary virtue is its simplicity; it is supposed to give us a parsimonious account of when an individual merits responsibility for the actions of others and when she does not. Kutz’s theory of participatory intentions (absent any further stipulations) is quite parsimonious. However, once we start adding further stipulations to the theory on an ad hoc basis, whether in the form of differing types of responsibility or plan-meshing requirements, the theory becomes less parsimonious. A minimalist theory with added stipulations is not really so minimalist after all. But if this is the case, then these amended minimalist intention theories are really just another form of complicated or sophisticated intention theory, such as the theories proposed by Tuomela and Mäkelä (2016) and S. Miller (2001) or the reduced luck authorization theory from chapter one. If a minimalist intention theory loses its parsimony, then conceptual simplicity is no longer a good reason to choose it over its competitors.

Because minimalist intention theories struggle to avoid disproportionate ascriptions of fault to marginal participants, we have reason to favor more sophisticated intention theories that do not run into similar problems. Among sophisticated intention theories, I suggest that we still have reason to prefer an authorization account over the others. To see why, let us take a closer look at the arguments that proponents of sophisticated intention theories put forward.

**Intentions, Intuitions and the Complicity Principle**

I argued in chapter one that an authorization theory of distributing collective responsibility accounts for how one individual becomes at fault for the actions of another. One individual shares blame for the actions of another if she authorizes the other person to act on her
behalf. By entering into an authorization relationship, an agent extends her autonomous
capacities through a voluntary agreement. She becomes the vicarious author of an outcome
through her agreement with another agent. Because the agent has willingly given her
authorization to another, it is not wrong or unfair to hold her responsible for the actions another
person carries out on her behalf.

Authorization, much like consent, is a concept that is acceptable to the strictest
methodological individualist. The viability of an authorization relationship between two or more
individuals does not require a collective mind, collective will or any other strong metaphysical
commitments. The transformative moral character of an authorization agreement also does not
require any collectivist metaphysical commitments. When A authorizes B to commit an injustice,
we blame A for B’s actions not because B literally becomes an extension of A’s body, but
because A produces the outcome through B’s agency by way of agreement. A’s fault is a
function of her agreement with B to bring about a certain state of affairs. As long as the
methodological individualist acknowledges that agreements (such as consent or authorization)
can be morally transformative, then we have a way to explain how one person can be at fault for
the actions of another.

An authorization theory of distributing collective responsibility therefore relies only on a
principle of individual responsibility. Individuals are at fault only for what they personally do or
omit to do, absent any excusing or justifying conditions. When individuals personally carry out
wrongdoing, they are typically to blame for their own actions and the reasonably foreseeable
consequences of those actions. When individuals form an agreement to have others carry out
wrongdoing on their behalf, they are typically at fault for the agreement they have made and the
reasonably foreseeable consequences of that agreement. Because the actions of an agent acting
on one’s behalf are part of the reasonably foreseeable consequences of the agreement, one becomes at fault for the wrongdoing the agent commits on one’s behalf, assuming one has properly authorized the agent to act.

But (minimalist and sophisticated) intention theorists as a group deny that consent or authorization agreements are a necessary component of distributing collective responsibility (see Kutz, 2000, p. 226; Sadler, 2006, p. 141; Mellema, 2016, p. 142). Their project is to demonstrate that the distribution of collective responsibility is a function of intention states alone, not any form of agreement. Without an appeal to authorization agreements, intention theorists will need a different strategy to account for how one person becomes to blame for the actions of another. Going by the principle of individual responsibility alone, it would not be clear why we should blame one group member for the actions another group member takes toward a shared goal, if we cannot refer to the agreements that they have made with each other.

The clearest justificatory strategy open to intention theorists is to argue that the rational structure of a collective, participatory or we-intention is such that, when instantiated by an individual, the individual acquires individual responsibility for collective wrongdoing. An individual intention to commit wrongdoing typically entails personal responsibility for that wrongdoing, so if collective or we-intentions to participate in a collective project prove to be structurally similar to individual intentions to commit wrongdoing, then perhaps collective or we-intentions are sufficient to normatively justify ascriptions of collective responsibility to individuals.

One could plausibly advance this argument in defense of Kutz’s theory of participatory intentions, as his participatory intentions are structurally similar to individual intentions to act. But as I argued earlier, we have normative reasons to reject Kutz’s minimalist account. Kutz’s
account requires extra ad hoc stipulations to ward off charges of unfairness, so neither parsimony nor conceptual simplicity are a reason to choose it over its competitors. So we must see if more sophisticated theories of collective intentions or collective ends are structurally similar to individual intentions to act. Unfortunately, by their very nature sophisticated theories of collective we-intentions are unlike individual intentions to act. Recall that Tuomela and Mäkelä (2016) interpret a collective intention as a plurality of intention and belief states, so they are decisively not like individual intentions to perform some specified act, which do not require such robust belief states about the actions of others. Because of this asymmetry, more argumentation is necessary to demonstrate that the sort of collective intentions Tuomela and Mäkelä propose entail personal responsibility for collective wrongdoing. Simply denoting a handful of mental states a “collective intention” does not give a compelling reason to hold some individuals responsible for contributions toward a collective end made by others, for it is not clear that collective intentions so described entail any meaningful normative considerations.7

Another way one could justify an intention theory is to postulate that when individuals participate in a collective endeavor with a shared aim, their wills enter into a special, linking relationship. Kutz (2000a) adopts this strategy at one point, arguing: “If a set of agents’ participatory intentions overlap, then the will of each is represented in what each other does qua group member, as well as what they do together. The logical overlap permits us to say that they manifest their attitudes through one another’s actions” (pp. 141-142). In other words, it is right to hold a person responsible for the actions of another agent if another agent’s actions manifest her will. Another agent’s actions manifest a person’s will when both instantiate a participatory

7 As I noted earlier, many theorists reject the thesis that a collective or we-intention entails any sort of personal responsibility for collective wrongdoing.
intention that aims at contributing to the same collective goal or collectively produced outcome. If another person’s actions manifest my will, then others have a reason to see that person’s actions as sufficiently similar to my own. If a collective, participatory or we-intention suffices to manifest one’s will in the actions of another, we have a compelling rationale for why one person may rightly be held responsible for the actions of another.

The normative leap from individual to collective responsibility therefore hangs on the extent to which two individuals’ wills might overlap by aiming at producing the same states of affairs. One’s will is no more than a collection of psychological states that aim at realizing some state of affairs. So the wills of participants in a collective action cannot literally overlap. What Kutz means by logically overlapping wills must therefore be a metaphorical claim about the normative relationships of two persons’ intention states. But this metaphor is not so clear. Two individuals can believe the same proposition without having their propositional attitudes enter into an overlapping, representing or manifesting relationship. If both you and I independently intend to participate in raising global temperatures by emitting more carbon, does this bring our wills into a manifesting relationship, even if we have no knowledge of each others’ contributions? I see no reason why the character of our wills would go through this transformation. It seems to me that if we wish to hold some individuals responsible for the contributions made by others, it is unwise to hinge the entire theory on an obscure metaphor that other theorists could simply reject. We must therefore look for a more compelling rationale.

Recall that Kutz also defends his theory of distributing collective responsibility by appeal to a normative principle: the complicity principle. Kutz’s formulation of the complicity

---

8 I focus on manifestation instead of representation to avoid wading into discussions about the meaning of representation; see Pitkin (1967). Alternatively, we could also say that my actions express or embody my will.
principle states that “I am accountable for what others do when I intentionally participate in the wrong they do or harm they cause” (2000a, p. 122). The complicity principle, when combined with a principle of individual responsibility, accounts for how one individual might become to blame for the actions of another. According to the complicity principle, once one participates in wrongdoing, one becomes complicit in wrongdoing. Participants can therefore rightly be held to blame for the actions of their fellow group members.

Armed with the complicity principle, the intention theorist of collective responsibility can justify distributing collective responsibility. It is acceptable to hold some individuals responsible for the actions of others because the complicity principle states that anyone who intentionally participates in another’s wrongdoing becomes complicit in that wrongdoing. We therefore need not make any reference to authorization or agreements. The complicity principle alone is sufficient to account for how we expand the principle of individual responsibility to include collective responsibility for the actions of others.

The success of this strategy rests on the extent to which Kutz can justify the complicity principle. He gives two arguments in defense of the complicity principle. First, Kutz defends the viability of the complicity principle by stipulating that it is “well-grounded in our institutions, ethical practices, and psychologies” (2000a, p. 122). Second, Kutz argues that when we take up a “second-person perspective” toward the suffering of those who have been harmed by collective wrongdoing, we are compelled to adopt the complicity principle (p. 123). I argue that neither strategy succeeds.

---

9 It is, after all, a controversial principle; see Gardner (2004, p. 829) and Gilbert (2002, pp. 182-183).
Consider first the fact that the complicity principle is “well-grounded in our intuitions, ethical practices, and psychologies” (p. 122). This strategy of defending the complicity principle on the grounds of our most personal attitudes and intuitions can succeed only if the aforementioned “we” actually exists and has come to some form of consensus about the viability of the complicity principle (or any other related principle of distributing collective responsibility). If it is indeed true that “we” rely on intuitions about the complicity principle in our everyday ethical practices, then this would be prima facie support for the complicity principle, assuming that we are willing to grant some justificatory power to our strongest intuitions.

The problem is that there are theorists (me included) who do not find the complicity principle intuitive. It is not firmly grounded in my emotions, intuitions and personal deliberations. It appears to me to be a post hoc rationalization of some of our nastier tribal instincts. H.D. Lewis (1948) does not find the complicity principle convincing. Neither does Velasquez (1983; 2003), Narveson (2002) or any other strong methodological individualist. So, to say that “we” find the principle of complicity intuitive is disingenuous. Some individuals find the principle to be highly controversial. To try to establish the complicity principle on the basis of controversial intuitions therefore begs the question against theorists who do not share those intuitions.

I will not rule out the possibility that theorists may establish core theoretical principles at least partially on the strength of intuition. But when attempting to resolve a theoretical dispute, we ought only to rely on intuitions that all or most parties of the dispute can reasonably share. To attempt to resolve the dispute by appeal to private intuitions that are not shared by some parties
is to settle the argument by begging the *key* question, and this is a highly objectionable strategy of argumentation.

Kutz also gives a competing-perspectives defense of the complicity principle. He argues that in cases of collective wrongdoing, there are three perspectives that we may adopt in response. First, there is the first-person perspective of the agent who participates in collective wrongdoing. From the agent’s perspective, she should not be blamed too harshly, because she was merely one participant out of many.\(^{10}\) We might reach a similar conclusion from an objective, third-person standpoint. From a third-person standpoint, we distribute blame in proportion to the magnitude of causal contribution, which for many participants in a large collective endeavor could be quite small (Kutz 2000a, p. 122). However, when we adopt a second-person standpoint, our perspective changes. From the second-person standpoint, we consider other moral agents not as causal objects but as autonomous agents worthy of dignity and respect (see Darwall, 2006). Kutz argues that, from the second-person perspective, we realize that the victim’s experience is “dominated by the fact of suffering” (2000a, p. 123).

Kutz argues that we need to reconcile reactive judgments made from each of the three perspectives. While the first- and third-person perspectives favor an individualist moral theory, Kutz argues that the second-person perspective provides warrant for the complicity principle (p. 123). Kutz concludes that the best way to balance the three perspectives is to weaken our traditionally individualist moral principles to make room for the complicity principle.

But a lot more work needs to be done to show that the complicity principle is a necessary component of balancing the three perspectives. How we form reactive judgments from the

\(^{10}\) In cases of overdetermination, Kutz (2000, p. 122) argues that participants in collective wrongdoing may play no clear causal role whatsoever. For an opposing view, see Petersson (2013).
second-person perspective is colored by the values we hold and the moral concepts we are
disposed to employ. A person could very well adopt a second-person perspective toward victims
of collective wrongdoing and still resolutely believe that the complicity principle goes too far.
She could consistently say: “were I in the victim’s place, I would only want responsibility$_f$ to fall
on the planners or leaders of such wrongdoing.” So there is reason to think that, even after ideal
deliberation and reflection, many may still question the necessity of the complicity principle, as
Kutz states it. It is therefore not at all obvious that the complicity principle is the best way to
balance judgments from the first-, second- and third-person perspectives.

We therefore have good reason to question the validity of the complicity principle and
other prominent defenses of intention theories of distributing collective responsibility$_f$.
Conversely, an authorization theory of distributing collective responsibility$_f$ does not rely on the
complicity principle. It relies only on a principle of individual responsibility$_f$, one which
proponents of other theories of distributing collective responsibility$_f$ accept as well. To reject the
principle of individual responsibility$_f$ would be, in effect, to dismiss the entire concept of moral
responsibility$_f$ as untenable (see Pereboom, 2001). So, while an authorization theory of collective
responsibility$_f$ relies on a principle of individual responsibility$_f$ shared by theorists of collective
responsibility$_f$, an intention theory either rests on obscure metaphors or a controversial
complicity principle that can only be established by begging the key question against critics.
These considerations give us a reason to prefer an authorization theory of distributing collective
responsibility$_f$ over competing intention theories.

**Conclusion**

Intention theories of distributing collective responsibility$_f$ are on the right track, but the
theories I have considered display shortcomings. Minimalist intention theories are conceptually
parsimonious, but the simpler the intention theory, the more it ascribes collective fault far too liberally to integral contributors and marginal participants alike. I have argued that assigning collective fault to marginal participants is disproportionate and therefore violates norms of fairness. This provides philosophers a normative reason to avoid minimalist intention theories of distributing collective responsibility. Conceptual parsimony is not worth the moral cost.

The deficiencies of minimalist intention theories may compel some to look more closely at conceptually complicated or sophisticated intention theories. By introducing qualifications as to which group members qualify for the distribution of collective responsibility, sophisticated intention theories avoid assigning blame too liberally. Sophisticated intention theories therefore do not violate principles of fairness in ascribing fault.

But more complicated and sophisticated intention theories are also subject to a serious objection. Sophisticated intention theories rely on principles of collective responsibility distinct from the principle of individual responsibility. While the principle of individual responsibility is relatively uncontroversial, principles of collective responsibility, such as Kutz’s complicity principle, are controversial. Intention theorists who wish to rely on a principle of collective responsibility must therefore provide other theorists a compelling or non-question begging reason to accept such a principle. So far, intention theorists have not produced such a reason.

These concerns speak in favor of an authorization theory of distributing collective moral responsibility. An authorization theory avoids ascribing disproportionate fault to marginal participants in collective wrongdoing and offers a justification for ascribing responsibility to some participants for the actions of others. We should therefore prefer an authorization theory to the prominent intention theories in the literature.
I have so far defended an authorization theory against some competing models of collective responsibility. These theories stipulate that the distribution of collective responsibility is a function of an intentional relationship to some actions or outcomes. There are also theories that distribute collective responsibility (both forward- and backward-looking) by appeal to our social and political ties. We can call this family of theories “social connection” theories of distributing collective responsibility. I now turn to a pair of prominent social connection theories of distributing collective responsibility to argue that they too face conceptual and normative objections.
CHAPTER THREE

BLAMELESS PARTICIPATION IN STRUCTURAL INJUSTICE

Discussions of collective and social responsibility extend beyond the Anglo-American philosophical tradition. While the thinkers of chapter two stand on the shoulders of H.L.A. Hart, Elizabeth Anscombe, Donald Davidson and others, in chapters three and four we consider theories of collective responsibility (both forward-looking and backward-looking) whose roots extend to the other side of the English Channel. A distinctive tradition of commentary on the problem of distributing collective responsibility emerges from the existentialism and phenomenology of thinkers such as Karl Jaspers and Hannah Arendt. Instead of grounding the distribution of collective responsibility in discrete intentions or intentional states, thinkers indebted to the continental tradition argue that the distribution of collective responsibility is a function of our social relationships. I will follow Iris Marion Young in denoting these theories “social connection” theories of distributing responsibility (2011, p. 96).

In this chapter, I focus on the theory of responsibility Young proposes in Responsibility for Justice (2011). She develops her social connection theory of responsibility in response to a form of wrong she calls a “structural injustice” (p. 52). A structural injustice occurs when members participating in one or more scheme(s) of social coordination act blamelessly, but the schemes, in conjunction with accepted norms and background conditions, systematically prevents some individuals from fulfilling their basic human rights or bettering their situation.

\footnote{Portions of this chapter have been reworked and published as Atenasio (Forthcoming 2019).}
Young argues that these consequences are a form of injustice. Young stipulates that a structural injustice is a legitimate form of wrong, but claims that it is not a wrong traceable to the personal misconduct of any one participant in any social scheme. For this reason, she argues that it is not right to blame any individual person for a structural injustice. Rather, Young proposes that all participants in a structural injustice have a forward-looking obligation to organize, vote, protest and advocate for social change. Participation in a structural injustice entails only a forward-looking responsibility to do what one can to remediate the harms caused by the scheme of coordination.

There is a sense in which the ends of an authorization theory of distributing collective responsibility as fault and a social connection theory of forward-looking responsibility diverge. It would be perfectly consistent to hold that backward-looking fault for the actions of others is always a function of authorization or command and that individuals accrue forward-looking moral obligations because of their social connections.

But I would like to cast doubt on whether a social connection theory of forward-looking responsibility is necessary. I argue that Young overestimates the extent to which a structural injustice may result from blameless participation. If we disambiguate what Young means by structural injustice, we find that genuine examples of a structural injustice involve substantial blameworthy participation. Those that do not involve blameworthy participation are not rightfully categorized as structural harms or injustices. Either way, we can generate sufficient forward-looking obligations either from a theory of fault or from traditional theories of beneficence or distributive justice.
This chapter proceeds as follows. First, to provide some context, I outline a brief intellectual history of theories of forward-looking collective or social responsibility as they emerge in the phenomenological and existentialist traditions. I say a bit about Karl Jaspers’s theory of guilt and Hannah Arendt’s theory of political responsibility. I then articulate Iris Marion Young’s social connection theory of responsibility. I rehearse some common criticisms of her account.

Next, I present my primary argument: that Young’s examples of a structural injustice entail substantial blameworthy participation. Any purported structural injustice that involves only blameless participation is not an example of structural wrongdoing. Rather, the wrong lies elsewhere, either in another unjust structure or society’s unwillingness to adequately distribute resources, benefits or burdens. Once we make this distinction, the need for a social connection theory of collective forward-looking responsibility loses its urgency. Remedial obligations for a structural injustice accrue to those who perpetuate and profit off of unjust or unfair structural relationships. If a fair and efficient scheme of social coordination merely fails to meet the needs of some community members, there is no structural injustice, but a failure by the community to discharge obligations of beneficence or distributive justice. In this case, remedial obligations accrue to individuals because they have been insufficiently beneficent or have ignored their duty to support just institutions. I conclude by responding to a number of objections to my criticism of Young’s theory of responsibility.

**Jaspers and Arendt on Political Responsibility**

Before turning to Young’s contemporary social connection account of collective forward-looking responsibility, let us begin with a brief survey of two influential works on the concept of
I first canvass Jaspers’s *The Question of German Guilt*. Jaspers penned this monograph in 1947, not long after the defeat of Hitler and the National Socialist party. His goal was to untangle the many claims of guilt and responsibility made in the wake of shocking revelations about the magnitude of Nazi malevolence during the war. Jaspers had to wrestle with two competing sentiments: that Germany was responsible for committing some of the cruelest barbarities in recent memory and that the majority of these crimes were primarily carried out by a minority of German soldiers, generals and officials.

To capture the complexity of German guilt for WWII, Jaspers proposes four different conceptions of guilt: criminal guilt, political guilt, moral guilt and metaphysical guilt. For Jaspers, criminal guilt is always juridical; it is legal responsibility for wrongdoing in the eyes of the law (2001, p. 25). When one performs an action that is legally prohibited, one is criminally guilty. Criminal guilt is primarily the concern of police and the court system. It is properly answered by the governing officials with various forms of punishment (p. 30).

Political guilt is guilt due to the actions of one’s state and elected officials (p. 25). As a member of a state or nation, one must bear the consequences of state actions (p. 25). The reason for this is that “everyone is co-responsible for the way he is governed” (p. 25). However, Jaspers insists that political guilt entails only forward-looking responsibility to rectify the harms caused by one’s state (p. 30). As a citizen of the United States, for instance, one is responsible for alleviating any wrongful harms caused by the American military or other governmental agencies.

---

2 For a discussion of the concept of responsibility in the continental tradition more generally, see Raffoul (2010).

3 For a competing perspective, see Goldhagen (1996).

4 Jaspers uses the word “guilt” (*Schuld*) to mean both backward-looking fault and forward-looking responsibility or obligation.
While one cannot rightfully be blamed for the wrongdoing of one’s governmental officials, one has an obligation to make good on the unjustified harm they cause.

Jaspers defines moral guilt as guilt for one’s intentional actions (p. 25). When one voluntarily performs a wrongful intentional action, one is morally guilty for one’s action. Jaspers insists that moral guilt is not alleviated when one is ordered to do wrong. He claims that actions performed by an individual under orders are still the individual’s intentional actions (pp. 25-6). The proper response to moral guilt is individual penance and renewal; one ought to feel sorrow over one’s transgressions and feel compelled to correct one’s behavior in the future (p. 30).

The theory of guilt that causes the most controversy (and confusion) is Jaspers’ notion of metaphysical guilt. Jaspers defines metaphysical guilt as a feeling that one is “co-responsible for every wrong and every injustice in the world” (p. 26). Every time that one fails to do the utmost in one’s capacities to prevent another’s wrongdoing, one is metaphysically guilty of that wrong (p. 26). Metaphysical guilt is a form of indelible moral stain (p. 26). It results from one’s inability to manifest unconditional love toward every one of one’s fellow human beings (p. 27). It is shame at one’s inability to be perfectly loving and protective of others in all of one’s personal conduct.

My interest in Jaspers’s work is his theory of political guilt (or obligation), as it is the conceptual progenitor of social connection theories of distributing forward-looking collective responsibility. However, in the Question of German Guilt, Jaspers offers no extended defense of his theory of political guilt. He claims that “it clearly makes sense to hold all citizens of a country liable for the results of actions taken by their state” (p. 33). Jaspers never explicitly justifies this claim. He makes some suggestive remarks: Jaspers claims, for instance, that “every
citizen is jointly liable for the doings and affected by the sufferings of his own state” (p. 46). He elsewhere states that political liability arises because the citizen “lives as a link” in a political chain with his fellow citizens “and cannot escape from their influence even if he was in opposition” (p. 70). These cursory remarks begin to sketch a theory of forward-looking political responsibility. But much is still unclear, for Jaspers does not elaborate on how the fact of our shared fate as fellow citizens grounds a general political obligation to rectify the wrongdoing of our government officials.5

For further elaboration, we can turn to Hannah Arendt, who develops Jaspers’s notion of political responsibility in a conference paper on distributing collective responsibility. Arendt agrees with much of Jaspers’s account of political, forward-looking responsibility. She also believes that responsibility for one’s political connections only entails forward-looking responsibilities or obligations (Arendt, 1987, p. 43). In her mind, claims of collective responsibility should always be grounded in political association; they are never appropriate for legal or moral contexts (p. 44). These claims of forward-looking responsibility for the actions of one’s state hold whether or not one’s government is properly representative (p. 45).

Arendt justifies her theory of political obligation with a few considerations. First, Arendt argues that citizens of a nation reap the rewards of its wrongdoing (p. 45). Because individuals receive benefits from their state’s misdeeds, individual beneficiaries are liable to use those benefits to ameliorate any unjust damages. This is not equivalent to Locke’s claim that individuals consent to the state’s rule by voluntarily accepting benefits (Locke, 1988, II.120). Rather, Arendt argues that receiving benefits from an unjust source is sufficient to ground a

5 For an attempt to reconstruct a justification from Jaspers’s cursory remarks, see Schaap (2001, pp. 750-1).
liability to make restitution for that wrongdoing. It is a form of unjust enrichment, even if it is involuntary unjust enrichment.\(^6\)

Second, Arendt argues that “every government assumes responsibility for the deeds and misdeeds of its predecessors and every nation for the deeds and misdeeds of the past” (Arendt, 1987, p. 45). Even a revolutionary government maintains a form of continuity with the regimes that came before it (p. 45). Therefore, as a matter of conceptual necessity, every government inherits a relationship to the deeds and misdeeds of its predecessors. Because a government acts to further one’s interests, one also, in a way, enters into a relationship to the deeds and misdeeds of past governing regimes. This relationship is not enough to ground fault or blame for a government’s misconduct, but it is sufficient for liability or forward-looking obligations to repair damages, Arendt believes (p. 45).

Finally, Arendt argues that our political obligation to rectify governmental wrongdoing is grounded in our nature as social creatures. The burden of accepting responsibility for the actions of our institutions is one of the prices that we pay for engaging in social and institutional action (p. 50). In the age of the nation-state, humans need institutions at all levels of subsidiarity, from local to national. These institutions are not only a necessary instrument for us to achieve our personal aims or preferences, they are a form of common human good. A life without the structured sociality of political institutions would be an imperfect or lesser form of life. With the privilege of participation in the common good comes forward-looking responsibility as well. Arendt believes that we ought to ensure the integrity of our public institutions, and this requires

\(^6\) Nozick famously questions whether there is anything morally relevant about the involuntary reception of benefits in (1974, p. 93).
us to assume responsibility for the consequences of institutional actions, even if we are not personally the authors of those consequences.

**Participation in Structural Injustice**

With Jaspers’s and Arendt’s theory of political responsibility in mind, we now turn to Iris Marion Young, who defends a similar account of forward-looking responsibility in *Responsibility for Justice* (2011). Young aims to build on the framework of political responsibility that Arendt develops in (1987). She agrees with Arendt that guilt or fault is not appropriate when making judgments about countries or societies (Young, 2011, p. 77). If every member of a nation is equally as guilty for a nation’s wrongdoing as every other, guilt or fault becomes a practically uninteresting concept. If guilt picks out everyone, it just as well picks out no one. However, Young criticizes Arendt for offering too broad a conception of forward-looking political responsibility (p. 80). Young argues that mere political association is not sufficient to generate political obligations (p. 80). She worries that a theory based on political association alone amounts to a form of guilt by association (p. 80). Young insists that, if we look through some of Arendt’s other political writings, Arendt occasionally requires more than mere political association to assign political obligations; she sometimes requires acts and omissions as well (p. 80). Young then develops her own theory of forward-looking political responsibility based on this broader interpretation of Arendt’s political writings.

Young’s primary interest is wrongs or injustices that are not easily attributed to the individual conduct of any one person. For isolated, individual wrongdoing, our concepts of fault or guilt work well enough. However, some injustices occur because, while members of a population blamelessly pursue their own self-interest, the system in which they participate, in
conjunction with accepted norms and background conditions, produces unjust outcomes (p. 52). Young calls these systemic wrongs “structural injustices” (p. 52). A structural injustice is an “invisible hand” style of wrong where an aggregate of micro-actions results in macro-consequences that none of the participants individually intend (see Nozick, 1974, pp. 18-22). To give Young’s own example, we can imagine a system of landlords, renters, real estate agents, lenders and apartment brokers who all participate in the real estate market. However, even if all members participate fairly, there may still be some renters who suffer housing insecurity (p. 45). A person could lose her job, be evicted from her home and then be unable to find a rental because she cannot come up with three months of rent in advance (p. 46). Young believes that in such situations, the housing-insecure renter suffers an injustice even though none of the members participating in the real estate market have personally wronged her.

Young argues that existing theories of distributing collective responsibility (either backward-looking fault or forward-looking obligation) insufficiently address structural injustices. She specifically singles out Christopher Kutz’s theory of participatory intentions as one she deems lacking. One can recall from chapter two that, according to Kutz, all who participate intentionally in a collective venture bear backward-looking responsibility for that venture’s outcome (2000a, p. 122). A necessary component of participation is having a conception of one’s action as contributing to the collective end or goal (p. 82). Because participants in a structural injustice by definition do not desire or foresee the consequences of their participation, Young argues that Kutz’s theory of participatory intentions has little to say about responsibility for a structural injustice (Young, 2011, p. 103). A similar criticism could be
aimed at the authorization theory of backward-looking collective responsibility, I defend in chapter one.

To address structural injustices, Young proposes a “social connection” model of forward-looking responsibility. She argues that all who contribute by their actions to a structural injustice share forward-looking responsibility for any unjust outcomes that result (p. 96). For instance, any individual who participates in a real estate market that leaves some people housing-insecure bears forward-looking remedial responsibility for that unfortunate outcome. She argues that all those who participate in a structural justice have an obligation to rectify or transform the unjust structural process (p. 96). This means that participants have an obligation to organize, speak publically, vote and raise people’s consciousness about the structural injustice (p. 112). Young argues that these sorts of activities can only be discharged through collective action (p. 105). That is why the forward-looking responsibility we acquire through participation in a structural injustice is always a responsibility we share with others (p. 110).

A point of clarification is necessary. Young sometimes implies that causal implication in a structural injustice is sufficient for forward-looking political responsibility (pp. xx, 95-96). But such an interpretation cannot be correct. Imagine that I am out for a walk one night outside my apartment when I am robbed at gunpoint. The next day, the robbery makes the news and it has the effect of driving property values down in my neighborhood. Because my being robbed caused housing prices to go down, I am therefore in the causal chain of events that led to the price drop.

Now, while my person becomes causally implicated in the change in real estate prices, I am not implicated because of anything I actually did or omitted to do. But if causal implication is
sufficient for the acquisition of remedial responsibilities, then it would seem that I am responsible for the real estate market’s harms due to my causal implication. Young is clear that she does not desire such a result (pp. 86-92). Thus, participation in a structural injustice cannot mean mere causal implication. Rather, one participates in a structural injustice when one intentionally acts to further one’s interests through a system of cooperation that, because of accepted habits and norms, results in an invisible hand type of wrong (p. 105). One’s participation is intentional under some mundane and blameless description, but not intentional under the description of causing or aiming at a structural injustice.

Many theorists have found Young’s social connection theory of political responsibility to be a helpful moral framework for discussions of global development, decolonization and the environment (Johnson & Michaelis, 2013; Meyer & Roser, 2010). Despite its popularity, Young’s social connection theory has been subject to a number of objections. The first of these comes from the foreword that Martha Nussbaum wrote for Responsibility for Justice. Nussbaum argues that pulling apart backward-looking fault and forward-looking obligation is not as simple as Young assumes (Young, 2012, xxi). The problem is that once we assign a political obligation to a party to rectify some harm, the party’s noncompliance appears to be a faulty omission. If I have an obligation to repair your broken bicycle and I ignore this obligation, then I have done something wrong. If I have done nothing wrong in ignoring this obligation, then it is not clear what normative reason I would have for carrying out any of my forward-looking obligations.

Although Young does not implicitly address this point, she eventually permits some backward-looking fault to creep back into her social connection model. Young believes that if we ignore our political responsibilities, we can be “criticized” for our inaction (p. 144). What
exactly Young means by criticism is not entirely clear, but deserving criticism usually involves some kind of liability to accusation and harsh language. If this is the case, then it looks like a social connection model of political responsibility does rely upon a backward-looking blame or fault conception of responsibility; after all. One of the purported benefits of Young’s model, that it gets us away from fault and blame conceptions of responsibility; is perhaps not such a large advantage after all.

Another worry with Young’s social connection model of political responsibility is that Young never defends the claim that people who suffer the consequences of a structural injustice are wronged and not merely unfortunate (Reiman, 2012, p. 743). Consider the renter who has some bad luck and loses her housing security. Perhaps she is forced to sleep at a homeless shelter for a week before she can find another apartment. Let us suppose that the renter has not been irresponsible or careless. Even though she works full-time at a fast food restaurant, because of rising rental costs, she has trouble securing a minimally decent apartment for her and her children.

Has the renter necessarily been wronged or is she merely unfortunate? Young claims that the renter’s situation is not merely misfortune, but that she has been done an injustice. However, Young never justifies or defends this claim beyond a quick aside that “most people would agree” that an injustice has been done to the renter (Young, 2011, p. 47). This point is important, because if the renter’s situation is not unjust but merely unfortunate, then it does not make sense to speak of duties of remediation, recompense or restitution (as opposed to general obligations of beneficence).
The problem is that many people will likely disagree with Young about the nature of a structural injustice. Political libertarians would claim that the renter who becomes housing insecure is unlucky, but that she suffers no clear injustice. Appeal to broad agreement about the nature of structural injustice will not succeed as an argumentative strategy here if no such agreement actually exists. Defenders of a social connection model of forward-looking responsibility will therefore need to give a more convincing defense of the proposition that structural injustices are legitimate injustices and not simply tragic misfortunes.

Another concern one might raise against a social connection theory of political responsibility is that it underestimates the extent to which fault can be ascribed for systemic harms. Young argues that it is difficult to pick out individual wrongdoers in many structural injustices when everyone’s contributions are overdetermined (p. 101). However, this may not be the case. For one, if we do not look for intentional wrongdoing, but instead try to determine how risky each individual’s actions were, we may find that many people who are legally blameless nonetheless acted in ways that unreasonably raised the risk of harm for others. For instance, when wealthy homeowners flock to a single region, they have the effect of driving housing prices up. This can raise apartment rental prices, which places a tremendous financial burden on the working poor. The wealthy homeowners who congregate together have acted within their legal rights. However, it is not outrageous to say that in preferring to live in a single fashionable location with other wealthy homeowners, they have engaged in behavior that raises the risk that poorer people will be priced out of their homes.7

7 If this example proves too controversial, we can generate other instances of individuals placing burdens on third parties through their self-regarding behavior. If a host of developers all develop property in one small region, and
Further, even if causal contribution to harm is indirect, that may not necessarily make it any less faulty or blameworthy (Barry and Ferracioli, 2013, p. 255). It is possible that a great deal of the global harms supposedly caused by “invisible hand” style structural injustices are actually the result of fiscal policies imposed on weaker countries by powerful countries (Pogge, 2008). For instance, sweatshops with abusive labor policies are not necessarily the product of chance or innocent business arrangements. People agree upon fiscal policy and they choose to operate their organizations one way instead of another. There is also reason to believe that many recent famines have been caused or greatly exacerbated by policy and human decision (Sen, 1981). So while causal contributions can often be complex and difficult to determine, once we demystify global structures by seeing them as a set of relations between concrete human agents, it may be possible to locate individuals who have made decisions that precipitate or exacerbate purported structural injustices.

Another objection is that a social connection theory of political responsibility does not distribute forward-looking obligations in a procedurally fair way (Neuhäuser, 2014, p. 244). If we foreground social connection as the most important factor in assigning responsibility for structural harm, then we necessarily place the same obligations to repair structural harms on the poor and wealthy alike, as both participate in structural injustices together. Poor people need to buy clothes just as rich people do, so both may purchase clothing made in sweatshops. In addition, even people who are exploited by global structures will have a responsibility to fix thereby raise the risk that flooding will occur in the community, then the developers all impose risks on members of the community without their consent.
those structures, as they are participants as well. Young’s model of political responsibility therefore threatens to place rather heavy burdens on the poor and powerless to organize, protest, speak publically, vote and raise people’s awareness. Given all the other burdens and stressors in the life of the poor and powerless, this seems immoral and unfair. There are more procedurally just ways of distributing the burden of bringing about social change and correcting the harms of structural injustices.

To address some of these concerns, Young introduces into her theory what she calls “parameters of reasoning” (p. 124). She believes we can use various guiding principles of thought to distribute obligations in a just way. For instance, Young argues that those with a large amount of power or privilege might bear increased responsibility for lobbying the government, voting and compensating victims of structural injustices (pp. 144-145). But this strategy begins to undermine her social connection theory of political responsibility. If we start assigning forward-looking responsibilities based on those who are most powerful and those who benefit the most from a structural injustice, then we are turning away from mere participation in a structural injustice as a meaningful category of assigning responsibilities. Instead, it looks like having power over the mechanisms of a structural injustice and benefitting tremendously from a structural injustice are the real generators of obligations. If the parameters of reasoning do all the work for distributing obligations, it is not clear why we would need to appeal to a social connection theory of distributing political responsibilities in the first place.9

---

8 Interestingly enough, Young sees this as an advantage of her view (2011, p. 113).

9 In addition, Young offers no clear rational justification for why we should accept her parameters of reason or see them as binding.
Disambiguating Structural Injustice

Young claims that a moral framework focused on individual fault cannot account for the peculiar wrong of structural injustice. She further argues that this shortcoming necessitates a social connection theory of forward-looking responsibility. The strength of Young’s argument rests on the extent to which examples of structural injustices due to blameless participation actually exist. If blameless participation commonly gives rise to structural injustices, then Young’s social connection theory will prove normatively valuable, assuming it can meet the objections rehearsed in the previous section. If, conversely, a structural injustice usually or always entails blameworthy participation, then the need for a social connection theory of forward-looking responsibility loses some of its urgency. If we can identify clear instances of wrongdoing that cause or perpetuate a structural injustice, we can place remedial obligations to rectify the structural injustice first and foremost on the wrongdoers.

I argue that, if we disambiguate what Young means by structural injustice, we will find that Young’s examples of a genuine structural injustice entail substantial individual wrongdoing. Purported examples of a structural injustice that do not involve wrongdoing are not really structural failings at all. Rather, they are widespread individual failures to act on duties of beneficence or distributive justice. Therefore, Young has given us no reason to think that examples of a structural injustice due to blameless participation actually exist. If this is the case, then a theory of participatory fault combined with obligations of distributive justice or beneficence is sufficient for assigning remedial obligations to rectify the injustices caused (or supposedly caused) by dispersed, structural forces.
Systematically Unfair Structures

One way Young describes a structural injustice is as a wrong that “exists when social processes put large groups of persons under systematic threat of domination” (2011, p. 52). A systematic threat of domination entails some power asymmetry that disadvantages or imposes unfair burdens on a person or class of persons (Stahl, 2017, p. 481). If a certain social structure supports, exacerbates, incentivizes or facilitates the continuation of a power asymmetry that severely disadvantages some class of individuals, without providing them with the means to better their situation, then there may be something unfair about the way that structure functions (Young, 2006, p. 114; 2011, p. 52). In Young’s preferred example, a renter cannot find adequate accommodations on the housing market because of accumulated background conditions that systematically disadvantage her. This suggests one possible interpretation of a structural injustice: an unfair cooperative scheme or schemes between parties seeking to advance their self-interest. Cooperative schemes of interaction ought to permit individuals to advance their mutual interests (Gauthier, 1986, pp. 83-85). If a cooperative scheme, in conjunction with accepted norms and background conditions, advances one party’s interests while placing the other party under a perpetual threat of domination and humiliation, then that is evidence that the scheme is prima facie unfair or operates against a background of unfair conditions.

There are two ways this could happen. A cooperative scheme could be unfairly manipulated by some of its participants. By “manipulation,” I mean the intentional utilization of threats, force, law or predatory pricing to gain a substantial bargaining advantage. For example, imagine that there are five wheat merchants who supply grain to a small village. The wheat merchants all compete to provide the best grain possible at the lowest price for the villagers.
However, one of the wheat merchants bribes the village constable to arrest the four other merchants and throw them into prison. The lone remaining merchant then drastically increases wheat prices to extract more profit from the villagers, who do not have any other good options. In this example, one wheat merchant uses force to prevent other suppliers from entering or remaining in the market. This permits the wheat merchant to gouge the villagers. This is unfair, both to her competitors, who are forced out of business and unjustly thrown in prison, and to her purchasers, who can no longer shop for competitive prices to feed themselves and their families.

We can also locate more subtle contemporary forms of manipulation that may not involve patently illegal activities. Let us return to Young’s example of an expensive urban housing market. As Young herself admits, landlords frequently collude and pressure each other to leave rent artificially high (2011, p. 51). In addition, members of urban middle and upper-middle class neighborhoods often petition and vote to prevent the building of affordable housing that would increase supply and drive down rent for low-income families (Cowen, 2017, pp. 6-7; Hsieh and Moretti, 2015). While residents may be within their legal rights to vote for restrictive zoning laws, their voting aims to block access to the market and maintain high housing prices. They therefore manipulate the market, and this seems somewhat unfair to low-income individuals who merely wish to find affordable accommodations.

Alternatively, one could argue that profiting off of certain social relations is unfair, because those relations are intrinsically exploitative or operate according to exploitative background conditions.¹⁰ Exploitation occurs when one party A extracts concessions out of

¹⁰ Libertarians typically argue that the notion of exploitation is incoherent; see Zwolinski (2007). They claim that purportedly exploitative agreements still require the assent of both parties. Because the poor and powerless voluntarily enter into employment contracts, businesses and business owners do nothing wrong by offering such jobs to people in poverty. But any libertarian who thinks exploitation is an incoherent concept is also going to deny
another party B that B would never agree to if B had a reasonable set of options to choose from (R. Miller, 2010, pp. 63-69). Party B only assents to the agreement out of desperation, poverty or severe need. If B must negotiate within a cooperative scheme from a position of severe vulnerability, and party A takes advantage of this severe vulnerability to extract unreasonable concessions, then party A exploits party B’s weakness.

Consider another of Young’s examples of a structural injustice: the international garment market. U.S. clothing manufacturers frequently offshore production to poor countries with depressed economies (2011, p.127). Many factory workers are young women who suffer abuse and sexual harassment from their bosses (p. 127). Working conditions are abominable, hours are long and bathroom breaks are rare (p. 127). Sick workers are fired and employees are not allowed to unionize. Workers who complain are threatened, beaten or killed while government officials look the other way (p. 127).

As described by Young, the international garment market involves patent wrongdoing. Threatening bodily harm to prevent employees from unionizing is a form of blameworthy labor market manipulation. Beating or killing workers to instill fear and facilitate compliance is wrong according to any number of moral theories. But even if we bracket these individual faults, it is still true that some U.S. companies rely on a foreign country’s poverty to extract concessions from the country’s citizens. Some companies likely could provide better wages and working conditions and still turn a lesser profit. However, they instead utilize their superior bargaining position to take advantage of the desperation of those in poverty. Those who take exploitation to

that an exploitative structural injustice is a form of wrongdoing. The fact that libertarians deny that an exploitative social structure is an injustice is not a threat to my analysis; rather, if libertarians are correct, it would support my claim that Young has not offered us anything resembling a structural injustice premised on blameless participation.
be an intrinsically unfair form of social relation will find something objectionable about the way these U.S. companies operate.

So we could interpret a structural injustice as a cooperative scheme that operates against a background of manipulated or exploitative conditions. But if we do, we will struggle to find an example of a structural injustice founded primarily on blameless participation. If it is an injustice to suffer harm as a result of exploitation or manipulation, then it is unjust to unfairly manipulate a scheme of cooperation or exploit others, absent any clear excusing or justifying conditions. Certainly gross examples of manipulation involve patent individual wrongdoing. The wheat merchant who bribes the constable acts both wrongly and illegally. But neither should we overlook the objectionable quality of more subtle forms of manipulation. Colluding with landlords to force rent prices artificially high and using zoning legislation to prevent the building of low-income housing are also morally objectionable. At the very least, if the lack of existence of low-cost housing in urban housing markets is an injustice, then using the law to prevent someone from building or finding low-cost housing would be prima facie wrong. Similarly, if exploitation is unjust, then actively exploiting another for one’s own gain is prima facie wrong.\(^\text{11}\)

A systematically unfair cooperative scheme of social coordination or cooperative scheme of social coordination that operates against unfair conditions therefore does not give us a plausible example of a structural injustice founded on blameless participation.\(^\text{12}\) Such a scheme does not become manipulated without some person or class of persons doing the manipulating.

---

\(^{11}\) I grant that not every participant in the garment industry is to blame for exploitation. But if exploitation is wrong, then I’m not clear why we would exempt executives and corporate board members from blame for choosing to offshore production to poor countries with exploitative working conditions.

\(^{12}\) For a further defense of this point, see Atenasio (Forthcoming 2019).
Similarly, a scheme of interaction cannot be exploitative if there is no one who intentionally enters into exploitative relations. If we want a plausible example of a structural injustice founded primarily on blameless participation, we will have to look elsewhere.

**Structures That Burden an Unconsenting Third Party**

Young writes that a structural injustice is a “moral wrong…[that] occurs as a consequence of many individuals and institutions acting to pursue their particular goals and interests, for the most part within the limits of accepted rules and norms” (2011, p. 52). We saw that a systematically unfair structure does not quite satisfy these criteria. But there is another way that a harm or systematic disadvantaging could emerge from parties seeking their self-interest through a scheme of coordination: a set of social relations could place unjustifiable costs or burdens on an unconsenting third party. For instance, Young offers the example of a financial crisis precipitated by certain trading behaviors (p. 63). Traders and speculators pursue profit by buying and shorting financial products, which may bring about a financial crisis that places burdens on other participants in the financial market. Although none of the traders desire or aim at a crisis, their activities end up imposing severe costs on others nonetheless.

For the sake of the argument, let us assume that there is nothing systematically unfair about the cooperative scheme in question, but that it still reliably produces some disadvantage or harm to a third party. There are four ways this could happen. First, a social scheme could involve a good, service or activity intentionally meant to harm a third party. Second, a social scheme

---

13 People need not understand their actions as exploitative to exploit others. I may believe that hiring a surrogate from a poor country is perfectly fair, even if I pay her very little. However, despite my beliefs about the surrogacy relationship, it could still be exploitative, for I could simply be mistaken. I do grant that a cooperative scheme could gain exploitative qualities through neglect or by accident. I will address this concern later.
could involve a good, service or activity that will plausibly harm a third party, even though participants do not desire or wish for the harm. Third, a social scheme could involve a good, service or activity that participants should reasonably have known would harm a third party. Finally, a social scheme could involve a good, service or activity that carries a negligible risk of harming a third party, but then the unlikely outcome just happens to obtain.

A system of social coordination could entail harm to a third party if individuals engage in an activity that intentionally aims at harming others. We can imagine an underground market where wealthy participants bid on and purchase the services of assassins who murder enemies and competitors. The market could be perfectly efficient and allow every participant a fair share of bargaining power. However, the market still facilitates harm to a third party because it allows individuals to intentionally trade a service that aims specifically at causing unjustified harm.

But some systems of social coordination still incentivize or cause a harmful outcome to a third party, even if no one desires or intends that harm. Consider the fossil fuel industry: it provides buyers with an opportunity to purchase fuel in order to operate vehicles, power plants and factories. However, there is good evidence that the burning of fossil fuels releases carbon dioxide into the atmosphere, which will continue to lead to rising global temperatures (Miller, 2010, pp. 88-92). These changes in global temperatures could lead to rising sea levels, shifting rainfall, higher ocean acidity and an increase in heat-related deaths (NASA, 2017).

Let us suppose that the fossil fuel industry is perfectly fair. All participants in the energy market have adequate bargaining power to purchase the fuel they need to live a decent life. No one supplier or purchaser dominates the market unfairly. However, although the market is fair, it could still be the case that the market incentivizes an increase in the production and use of fossil
fuels, which thereby causes an unreasonable or unjustifiable magnitude of harm to the environment that we all share. This harm will acutely afflict those who are poorest and most vulnerable, preventing them from fulfilling their rights and bettering their situation (Miller, 2010, pp. 98-99).

Some may also engage in activities that they do not realize will impose externalities on a third party. Suppose I purchase an order of pesticides to use on my lawn. Perhaps I do not know that the pesticides are harmful. I use the pesticides and they leach into the lake near my house, causing it to become unsuitable for wildlife or recreation. Even though I did not realize the risk, I still purchased a harmful product using a cooperative scheme that we may presume is fair and efficient.

It could be the case that I should have known that the product was harmful. Perhaps I purchased the pesticides from an illicit distributor, which should have given me a strong indication that there was something wrong with them. But there are examples when it is not reasonable to expect a harm to occur. Assume that the pesticides proved to be safe in all clinical trials. However, a trace element of a rare chemical in my soil combined with the pesticides to create a toxic substance that ended up killing wildlife and preventing others from utilizing the lake for recreation. It was perfectly reasonable for me to use the pesticides, given what was generally known about them. I was just really unlucky.

In all four of these examples, some form of harm or damage to a person’s body or interests materializes from a fair scheme of interaction. But we have not yet found a compelling example of a structural injustice founded on blameless participation. It is wrong to purchase goods and services that are intended to cause unjustified harm to others, such as murder-for-hire
contracts, because it is wrong to intentionally cause unjustified harm to others. We justifiably outlaw people from entering into such contracts. Similarly, if all parties know that their activities impose unreasonable burdens on a third party, then it is wrong to continue with those activities. If the use of fossil fuels is unreasonably or unjustifiably harmful to the poor and vulnerable, then buying, selling or using fossil fuels is prima facie blameworthy in some way. While we might want to exempt from blame those too poor to explore sustainable energy options, surely we would not want to exempt wealthy investors and corporations from blame for profiting handily off of the sale of fossil fuels, if they know those fuels to be unreasonably or unjustifiably harmful to the environment. If the environment has not degraded to the point that buying and selling fossil fuels is morally objectionable, then we could imagine a situation in which the environment has grown far worse. Eventually, we would have to concede that the continued buying, selling and profiting off of fossil fuels is wrong because it would place unreasonably heavy burdens on unconsenting third parties, and we would be justified in legally restricting it.

But neither is it much better to engage in a harmful activity in ignorance of its obvious risks. It is both morally and epistemically blameworthy to do no research whatsoever and then purchase and use pesticides from a patently illegitimate source. The law would not countenance a defense of ignorance in this example and neither should we. While an ignorant user of dangerous pesticides would not be to blame for intentionally causing harm, she would still be to blame for acting negligently or irresponsibly (see May, 1992, pp. 42-46 and Zimmerman, 1986). Negligent conduct, while not as serious as intentionally causing harm, is still blameworthy.

When a typically harmless good, service or activity happens to materialize some remote risk, then that is obviously bad, but it is not an example of a structural injustice. Remember that a
structural injustice entails some sort of systematic disadvantaging or harm to either one of the parties of the agreement or to a third party. But if a safe good, service or activity very rarely causes harm to a third party, then there is no possibility of a systematic disadvantaging. A systematic disadvantaging requires a process that reliably causes harm to, imposes burdens on or prevents the improvement of some individual or individuals. By its very nature one cannot reliably impose a fluke harm on others, so neither can it count as a plausible example of a structural injustice.

A cooperative scheme of social coordination that visits a harm on an unconsenting third party cannot fulfill Young’s criteria of a structural injustice founded primarily on blameless participation. Either individuals are to blame for intentionally engaging in an activity that reliably imposes burdens or harms on an unconsenting third party or the scheme of coordination is typically safe and does not reliably impose burdens or harms on anyone. Either way, we need to keep looking.

Structures that Inadequately Distribute Resources

Consider once again Young’s primary example of a structural injustice: a housing market in which landlords, homeowners, lenders, brokers, renters and real estate agents participate blamelessly. Assume that each participant has an equal share of bargaining power and the market operates optimally. However, no matter how well the housing market operates, there may still be a few people who, due to circumstances outside of their control, fail to procure affordable housing for months or years at a time. This example does not cleanly fall into either of the two previous categories. Rather, the housing market cannot adequately provide housing to everyone in the community, even if prices are fair. This leads us to a third possible interpretation of a
structural injustice: a cooperative scheme of social coordination that inadequately distributes resources to members of a population.

We need to be careful here, as when it comes to basic material welfare, it is difficult to isolate a single market or scheme of resource distribution as the cause of a person’s shortcomings. The housing market could be efficient and fair, but some may still be unable to purchase housing because they cannot obtain a livable wage due to a very unfair labor market. Or, if the market for food is unfair, that may prevent people from purchasing other goods and services that they need. So for the sake of the argument, we need to presume that the markets for labor and vital goods and services are all fair, yet also that some people still fail to meet their basic needs.

Although I doubt that such a system of markets actually exists, it is nonetheless a possibility (Gauthier, 1986, p. 103). However, were such a system of fair but insufficient markets to obtain, then neither would they present a compelling example of a structural injustice founded on blameless participation. They would not amount to a structural injustice at all. To rightfully attribute an injustice to a structure, we must show that it caused some harm or failed to cause a benefit that it ought to have caused (see Young, 2014, p. 104-105). But the markets, institutions and social norms in question are not facilitating exploitation or transactionally unfair economic relationships. Were the fair market systems to disappear, the participants would be made worse off. It is therefore unclear why the structures and coordinative schemes are to blame here, nor why market participants would acquire remedial obligations on account of their participation. To place obligations only on those who participate blamelessly and not on others seems arbitrary and unfair.
The salient moral failing in question is that the markets and norms that make up the structure do not provide everyone with everything that they need to lead a minimally decent life. But it is not clear why we ought to expect markets or norms to cover shortfalls due to bad luck or natural inequalities, especially when legal and political institutions tasked with such resource redistribution already exist. Markets typically function to exchange resources, not to ensure that all people fulfill their basic human rights or needs. If we ought not expect a system of markets, background conditions and social norms to fulfill everyone’s basic rights, then individuals blamelessly interacting through these structures are the wrong target of remedial obligations.

It could be that the single mother’s fellow citizens are failing to meet basic standards of distributive justice. Perhaps there is some distributive scheme that all rational agents would agree to while choosing principles of justice behind a veil of ignorance (Rawls, 1971, pp. 17-22). If the citizens do not collectively actualize the distributive scheme required by the ex ante agreement, then they wrong those who go housing- or food-insecure by failing to discharge their basic duty of justice. But this injustice stems from the omissions of individual citizens who fail to vote for and support a just scheme of distribution; it has nothing to do with systemic failures.

It might also be the case that the single mother’s fellow citizens fail to discharge duties of beneficence. Maybe we have an obligation, assuming that all of our needs are met, to make sure that our fellow citizens or human beings do not suffer starvation or homelessness. If so, then citizens act wrongly by failing to prevent housing and food insecurity in their community. Again, the moral failing here is not structural but individual. It results from omissions on the part of citizens who refuse to meet the needs of others, either through charity or political action; their actions are insufficiently beneficent or benevolent.
It is possible to imagine a fair scheme of social coordination that nevertheless leaves some people with inadequate resources. But it is incorrect to call this a structural failure. By stipulation, we assume the schemes of coordination that make up the structure are operating fairly. A fair cooperative system of resource distribution is still preferable to no system at all. The wrong, assuming there is one, must therefore lie either in the citizens’ failure to discharge basic duties of distributive justice or their failure to discharge duties of beneficence.

A scheme of coordination or resource distribution that fails to meet the needs of some members of the community is not a plausible example of a structural injustice. Rather, it is the product of individual failures to discharge duties of distributive justice or beneficence. Therefore, we have yet to locate a clear example of a structural injustice founded primarily on blameless participation, even though we have canvassed the possible interpretations of Young’s writings on the subject. It is thus incumbent on proponents of Young’s theory of social responsibility to locate an alternate interpretation of structural injustice that is plausibly founded on blameless participation, for Young provides us with no good reason to think that such a phenomenon exists. Otherwise, it looks like we can adequately distribute fault and remedial obligations with traditional individualist theories of blame and forward-looking responsibility.

**Objections**

I have argued that a social connection theory of forward-looking responsibility is not necessary to assign obligations for remediating structural harms and injustices. We can either assign remedial obligations to those who exploit others, unfairly manipulate the market or sell harmful products, or we can assign obligations to citizens by appeal to duties of beneficence or distributive justice. In both cases, no appeal to social connection is necessary.
I wish now to consider some potential objections to my argument.

Some might worry that duties of beneficence and distributive justice are normatively suspicious, whereas Young’s interactionist duties are not. But if positive duties of beneficence and distributive justice are normatively questionable, then forward-looking duties of political obligation that do not derive from an individual’s consent, authorization or patent wrongdoing are also normatively questionable. Were it the case that Young’s forward-looking political responsibilities were due to blameworthy actions on the part of participants, we could say that participants acquire remedial obligations because they violate negative duties not to harm. But by definition blameless participants in a structural injustice do not violate negative duties not to harm. Were they violating negative duties not to harm, they would act in a blameworthy manner. Most participants in a structural injustice are presumably justified in their participation.

So as long as proponents of a social connection theory of responsibility want to claim that a structural injustice entails largely blameless participation, they will have to admit that political obligations to remedy a structural injustice do not derive from an individual’s consent, authorization or wrongdoing. But if so, then obligations to remedy a structural injustice are equally as normatively questionable as duties of beneficence and distributive justice, which also do not derive from an individual’s wrongdoing, consent or authorization.\textsuperscript{14}

One could further object that a structural injustice might occur by accident or without any party’s knowledge.\textsuperscript{15} If so, it looks like we have a candidate for a structural injustice founded on

\textsuperscript{14} Obligations of distributive justice could be the result of hypothetical forms of consent or authorization; see Stark (2000) and Stilz (2011).

\textsuperscript{15} I thank Jonathan Spelman for raising this objection.
blameless participation, assuming that the parties were not negligently or willfully ignorant about the risks they imposed on others. For example, imagine that a large, socially conscious firm develops a cheap source of protein. They introduce their product to market, causing a drop in demand for more expensive forms of protein. Other firms quickly work to develop a competitive product. However, purely by accident, as each other firm introduces a competing product, the socially conscious firm just happens to drop the price of their product and sell at a loss until the competitor is put out of business. This is not done maliciously, but because of miscommunication among members of the socially conscious firm’s bureaucracy. After a few years of such accidental price-dropping, the socially conscious firm ends up merging with the failing competitors and becoming a monopoly. Purely by coincidence, the firm then raises prices on their product, causing some people who formerly relied on it to suffer malnutrition.

At first glance, it is not immediately clear that we should call accidents such as this an injustice. If the intentions of the socially conscious firm’s executives and board of directors are good, then perhaps the result is not an injustice but an unfortunate tragedy. If so, I struggle to see how we could claim that people who can no longer afford the firm’s source of protein have been treated unjustly. But more importantly, this example of a fluke harm visited upon a population cannot satisfy one of Young’s criteria of a structural injustice. Recall that Young requires that a structural injustice systematically place some population under threat of domination. By definition a fluke occurrence cannot systematically or reliably place a threat on anyone. So an accidental market tragedy is not a good example of a structural injustice. If a firm continues to “accidentally” pursue a monopoly over many years, we are probably justified in questioning whether they really have benevolent motives after all.
Now it is possible that the socially conscious firm’s market manipulation is not the result of an unlucky accident, but rather due to managerial incompetence. Managerial incompetence could reliably or systematically visit a harm on an unconsenting third party. But if so, then we would have good reason to blame the managers at the socially conscious firm. We can rightfully blame them for being negligent and not taking due precaution. So again, we struggle to find a structural injustice founded on blameless participation.

One way to potentially save Young’s account is to stipulate that a structural injustice is a normatively neutral harm that emerges from individuals blamelessly participating in some cooperative scheme of social coordination. The harm need not be unjustified or unreasonable, just some sort of damage to or limitation of a person’s interests. If so, then it is quite simple to generate examples of a structural injustice founded primarily on blameless participation. Individuals and companies impose small risks and costs on others all the time. Plenty of cooperative schemes entail that some people fail to satisfyingly advance their interests. So any cooperative scheme that imposes these justifiable harms, risks or costs on others would constitute a structural injustice.

There are many examples of justified harms emerging from blameless participation, but they will not do the normative work that Young needs them to do. Imagine that I and my arch-nemesis both apply for an assistant professor job at a research university. Both of us do a great job in the interview and are worthy candidates. Also, both of us desperately need the job to provide for our respective families. The competition is fierce, but my nemesis ends up getting the job and I am forced to work as an adjunct for the foreseeable future. Here we have an example of

---

16 I thank Audra Goodnight for this objection.
a scheme of interaction that has harmed my interests, even though all of us acted in a blameless fashion.

But it is questionable whether the resulting harm to my interests entails any normative consequences for my nemesis or the research university. Neither my nemesis nor the university have an obligation to provide me with decent work simply because they prevented me from getting the job. I had no right to the job, so no one incurs obligations by preventing me from obtaining it. Herein lies the problem with modifying Young’s account to include justified or non-justified damages or limitations to a person’s interests. If a harm is entirely justified or not in need of justification, then no one incurs any obligations to remedy that harm. So any example of a structural injustice that produces justified or justifiable harms will be normatively inert and won’t produce the sort of forward-looking obligations that Young requires.

The astute reader will notice that in criticizing Young’s conception of a structural injustice, I have covered only some examples of what people typically refer to as structural harm or wrongdoing. All of Young’s examples involve a structural injustice that emerges from cooperative systems of resource distribution (Young, 2006, p. 119), but structural and systemic harms may emerge from non-cooperative systems of interaction as well.\textsuperscript{17} One clear example of this is when racist policies or norms become entrenched in a system of law or government. People who are unfairly treated by the government are typically not willing participants in some unfair or inequitable scheme of cooperation. Rather, the unjust government or legal system confronts oppressed minorities as a form of coercive imposition.

\textsuperscript{17} In (2011), Parekh argues that customary or legal gender inequality counts as a form of structural injustice.
Therefore, some may argue there are structural injustices that emerge from non-cooperative systems due to blameless participation. But an analysis of non-cooperative systems will turn out much like the previous analysis of cooperative systems. First, some government oppression or injustice is the result of clear intentional wrongdoing. School segregation policies in the United States were not oversights or accidents but deliberately enacted policies bolstered by the ruling of Plessy v. Ferguson. Even as late as 1912, progressive Woodrow Wilson intentionally reversed the Federal government’s reintegration efforts (Ambrosius, 2007, p. 699). We need no invisible-hand theory of structural injustice to explain the American government’s racist policies before civil rights legislation, for they were purposely enacted by voters and government officials.

However, one generation could establish an unfair or oppressive government system that the following generations perpetuate. When enough generations have passed, some government officials may want to claim: “why blame me for the racist system? I wasn’t the one who set it up or voted for it.” Some may view this as a form of blameless participation in structural injustice, but I have my doubts. First, there is one clear locus of blame: the individuals who first established and set up the racist or oppressive government apparatuses. The problem is not that we do not know who to blame, it is that the individuals who deserve blame have all passed away. While this answer may be unsatisfying, it does clearly locate a set of individuals who are at fault.

In addition, individuals who continue to benefit from or perpetuate an oppressive or racist system deserve blame for benefitting from, supporting and perpetuating that system. Any individuals who knowingly benefit off of the oppressive or racist system despite having the
power to change the system are at fault for a form of unjust enrichment.\textsuperscript{18} Or, they act negligently or fail to take due precaution by leaving potentially harmful norms and laws in place (May, 1992, pp. 46-52). So there is good reason to think that those who help perpetuate systemic forms of racism are not innocent or blameless participants in a non-cooperative scheme.\textsuperscript{19}

But while it is difficult to imagine a system of racist norms and background conditions emerging from agents blamelessly pursuing their self-interest in good faith, the same may not be true for sexist norms and background conditions. Because of the capacity to get pregnant, some women may find themselves systematically disadvantaged on the labor market.\textsuperscript{20} This disadvantage could persist, even if we stipulate that the labor market is optimal and everyone acts blamelessly and bargains in good faith. If a woman must refrain from working for three months or more with the birth of each child, she may find it hard to compete for decent wages with others who do not have children. We could therefore imagine an optimal labor market that systematically disadvantages mothers, preventing them from improving their situation and fulfilling their basic human rights. Perhaps this is an example of a structural injustice that emerges due to blameless participation.

There are at least two ways we can think of fairness: transactional fairness and distributional fairness. Norms of transactional fairness govern the agreements we make with one

\textsuperscript{18} This is not equivalent to the claim that individuals acquire moral obligations by passively receiving benefits; see Nozick (1974, pp. 90-95). Rather, individuals who accept the benefits of a racist system are at fault for knowingly receiving those benefits from an unjust source, in much the same way that one is at fault for knowingly receiving stolen merchandise; see Cuoto (2017).

\textsuperscript{19} Hence Charles Mills’s use of scare quotes around the term “innocent” when discussing individuals who lack ill-will toward African Americans but who reproduce systemic racism nonetheless; see (2003, pp. 60-61).

\textsuperscript{20} I owe this objection to César Cabezas Gamarra.
another. If the parties are fully informed, the agreement violates no moral norms and no party takes advantage of any other, then typically we say the transaction is fair. Conversely, if one party substantially deceives or takes advantage of another, then the transaction is unfair.

Distributional fairness governs the allotment of basic goods, resources, benefits and burdens in society. If everyone receives what they are owed, then a society is distributionally fair. If some do not receive the basic goods or resources they are due or bear greater burdens than they ought to, then a society is distributionally unfair.

Whether or not an optimal labor market that disadvantages mothers counts as a structural injustice depends on whether it constitutes a violation of transactional or distributional fairness. If the labor market is transactionally unfair, then an optimal labor market that disadvantages mothers counts as a systematically unfair scheme of social coordination. If the unfairness results from a distributional shortfall or inequity, then the optimal labor market proves to be an example of society failing to meet the basic needs of some of its members.

If we take the unfairness experienced by mothers in the optimal labor market to be an example of transactional unfairness, then it is not the case that all companies participate in the labor market blamelessly. If it is transactionally unfair to refuse to hire women of childbearing age or to pay them a lower wage, then companies deserve blame for adopting such policies. If the unfairness is distributional, then the companies act blamelessly, but the injustice is attributable to society; it results from the failure of individual citizens to discharge duties of distributive justice. Whether we understand the optimal labor market that disadvantages mothers as an example of a systematically unfair scheme of coordination or an inadequate distributional scheme, we still do not have a structural injustice premised on blameless participation. Either the
scheme itself is unfair and some deserve blame for utilizing it, or the scheme is fair but the failure is attributable to society at large.\footnote{My own sense is that an optimal labor market that disadvantages mothers counts more as a failure of distributional fairness than transactional fairness. In the case of disability, the U.S. levies a payroll tax to fund a federal insurance program to provide income for individuals who are unable to engage in substantial gainful activities. A similar insurance program could likewise provide an income for women who cannot work for maternity reasons. But to tax either all companies or citizens is effectively to shift the burden for alleviating inequities due to disability or pregnancy from the individual company to society. It is to treat the injustice as a matter of distributional unfairness rather than transactional unfairness.}

One further objection might be that our best psychological literature indicates that claims of fault and blame sometimes make rectifying extensive harms more difficult (Darby and Branscombe, 2014). Blamed or faulted individuals often grow resentful and defensive. If it is the case that blame is the incorrect response to a structural injustice, even if patent wrongdoing is clear, then perhaps we have pragmatic reasons to ignore a theory of fault for participation in collective wrongdoing and instead use a social connection theory of responsibility. There is good evidence that often blame is far less effective than forgiveness and reconciliation (Darby and Branscombe, 2014, pp. 127-128). But we should make sure to keep the metaphysics of wrongdoing distinct from other normative considerations about how we ought to respond to wrongdoing. It could be the case that we should not punish some wrongdoing and instead focus on the positive project of repairing injuries and developing new mechanisms of accountability (Weitekamp, 1993). But if so, we do not efface the fault of participants in a structural injustice, we merely forebear to blame them. Forgiveness still entails that there is something to forgive. So it is no objection to my analysis that we have pragmatic reasons to occasionally overlook the faulty actions of others.
One final objection is that we can locate a structural injustice in a certain kind of resource shortfall. Imagine that multiple countries compete to fish in a shared body of water. Some countries find more success than others. Eventually, because of a high demand for fish, a few larger countries end up overfishing the body of water. The fish population dwindles and some of the fishing teams from the poorer countries can no longer catch enough fish to feed their citizens. Some people end up starving. Assuming that all fishing parties acted in a justified manner, the resulting starvation may constitute a structural injustice premised on blameless participation.

First we need to distinguish whether the wealthy countries extract the fish as a basic food source or as a luxury. If the wealthy countries overfish because they desire to make jewelry out of fish scales or because citizens demand a wasteful delicacy that only utilizes a small portion of the fish, then they are to blame for their profligacy in the face of need. If my neighbor, because of his poverty, lives off the berries in the public woods behind my house, I act wrongly by picking all the berries and turning them into a clothing dye, especially if I have plenty of other good ways to color my clothing. If there are scant resources to go around, it is not acceptable to waste those resources on amusements and trivialities. This principle follows from basic duties of beneficence: if we can help a struggling person without incurring unreasonable costs, then we ought to do so.

But imagine that no country wastes the fish. The overfishing results from the fact that all neighboring countries rely on fish for their diet. All countries act blamelessly, but due to limited

---

22 I thank James Cain for raising this objection.

23 This is a much weaker principle than the Lockean Proviso, which requires gatherers of resources to leave enough left over so that everyone may extract a fair share.
resources, some cannot acquire the fish that they need. If such a resource shortfall is an injustice, then we have located an injustice premised on blameless participation. But I’m not sure it is an injustice. Garret Hardin (1968) famously calls the lack of care for common resources a “Tragedy of the Commons.” The term tragedy indicates that it is not a wrong visited on anyone, but merely a misfortunate outcome based on the fact that no one polices the use of natural resources. Indeed, a resource shortfall seems to me a paradigm case of a moral tragedy, assuming no parties violate clear duties of beneficence or distributive justice. So while a resource shortfall might be unfortunate, it is not necessarily an injustice visited on any party.

**Conclusion**

Young defines a structural injustice as an unintended wrong that occurs on account of widespread blameless participation in a cooperative social scheme or schemes. But it is not clear that any actual scheme of social coordination (in combination with accepted norms and background conditions) satisfies these criteria. If a scheme is unfair or unreasonably harmful, then at least some class of individuals will justifiably be to blame for wrongdoing. We should not consider manipulating the market, exploiting others or engaging in activities that unreasonably burden third parties to be, as Young puts it, “within the limits of accepted rules and norms” (2011, p. 52). Conversely, if a scheme fails to adequately meet the needs of some community members, that may be an injustice, but it is not a structural injustice. The injustice or wrong, assuming it is an injustice, is a function of inadequate beneficence or a failure of the citizenry to discharge their basic duties of distributive justice.

---

24 If Beitz (1979, p. 141) is correct to say that the demands of distributive justice reach across borders, then most resource shortfalls do entail the failure of some wealthy countries to meet basic obligations of justice.
In an era of widespread industrialization and globalization, it may be difficult to know who, if anyone, we ought to blame for outcomes that anger or disappoint us. Many of our interactions are mediated through byzantine market structures and supply chains that disperse all over the globe. The idea of a structural injustice gives us one way to normatively criticize these systems. However, as I have argued, once we specify more precisely what it is we mean by structural injustice, it looks like we are dealing with several sorts of wrongs. Schemes of coordination can be used as tools of manipulation or exploitation to extract unreasonable concessions out of those with little bargaining power. Sometimes, they inadvertently place burdens on unconsenting third parties. Cooperative relationships may also fail to adequately distribute resources. If the schemes of coordination are otherwise fair, the problem lies not with the schemes but with a failure to discharge duties of beneficence or distributive justice. In either case, we should not be afraid to pierce the corporate veil and place some blame where it is due.

I have so far surveyed one popular theory of collective forward-looking responsibility that emerges from the continental tradition. I have argued that we do not need to look beyond an authorization account of collective responsibility (in conjunction with general duties of beneficence or distributive justice) to properly assign remedial obligations for structural wrongdoing. But as I promised at the beginning of the chapter, there is still one more contemporary theory of collective responsibility indebted to the traditions of existentialism and phenomenology: the negligent-attitudes model proposed by Larry May. May ties responsibility for the actions of others to personal negligence for cultivating harmful attitudes. In the next chapter I criticize May’s account and argue that we still have good reason to prefer an authorization account of distributing collective responsibility.
CHAPTER FOUR
NEGLIGENCE AND INTENTIONAL WRONGDOING

Like Iris Marion Young, Larry May (1992; May and Strikwerda, 1994) develops a distributive theory of collective responsibility by appeal to the continental traditions of phenomenology and existentialism. But despite emerging from similar roots, May’s position differs from Young’s social connection model. Whereas Young focuses our attention on the social structures in which we participate, May urges us to look primarily to the attitudes we cultivate. More specifically, May attempts to link collective or shared responsibility to harmful attitudes. He argues that group members who continue to negligently cultivate or harbor racist or sexist attitudes share responsibility for intentional wrongdoing carried out by their fellow racist or sexist group members. May calls this model of collective responsibility a social existentialist theory of responsibility (1992, p. 19).

May’s social existentialist theory is significantly more expansive than my own authorization account. Whereas an authorization account distributes responsibility only to those who authorize wrongdoing, May distributes at least some responsibility for collective wrongdoing to group or community members who simply harbor hateful attitudes. If May’s account is plausible, then an authorization account of distributing collective responsibility proves insufficient, as it fails to distribute responsibility for collective wrongdoing to those who only adopt hateful attitudes toward another person or group.
But I will argue that May’s account is not tenable. By using negligence and negligent attitudes to distribute shared responsibility, May calls for a substantial revision to what I will call the standard liberal view of how negligent actions relate to intentional wrongdoing. According to the standard liberal view, an unjustified harm intentionally caused by one party typically cancels out or overrides any normative connection between the untoward outcome and the negligent behavior of others. For example, if you negligently leave your pool uncovered and a perpetrator drowns a small child in your pool, according to the standard liberal view, it is the perpetrator who is responsible for the child’s death, even though you may have been irresponsible in leaving your pool uncovered.

May’s account requires us to deny the cancelling effect of intentional wrongdoing, and this leads to a number of unwelcome consequences. First, without the cancelling effect of intentional action, May’s account greatly increases the number of people who are responsible for any given wrongful outcome. This expansion threatens to weaken the functional efficacy and normative force of responsibility ascriptions. Second, it collapses the clear distinction between negligence and intentional action, thus weakening our ability to make fine-grained ascriptions of moral responsibility and legal liability. Third, May’s account suffers from a handful of line-drawing problems. He has trouble delineating both where a person’s community begins and ends, but also the threshold at which an attitude qualifies as hateful. These line drawing concerns are not easily dispelled with standard solutions. Finally, May’s account violates both norms of fairness introduced in chapter one: the moral separateness of persons and proportionality.

Therefore, we ought to resist May’s social existentialist theory of distributing collective responsibility and stick to the standard liberal view of how negligence relates to intentional action.
This chapter proceeds as follows: first, I articulate the concept of negligence and explain how it typically functions in moral and legal theory. Second, I present May’s social existentialist theory of responsibility, which attempts to link negligent attitudes to shared (or collective) responsibility. Next, I contrast May’s understanding of how negligence relates to intentional action with the standard liberal view as articulated by Hart and Honoré (1985). I then offer some objections to May’s account and argue that these objections give us good reason to reject May’s social existentialist theory of distributing collective responsibility and stick with the standard liberal view of how negligence relates to intentional action. I conclude by addressing several objections to my analysis.

**Negligence**

Western moral and legal theorists typically focus their efforts on theories of intentional wrongdoing. Often, an agent is at fault for an outcome if that agent intentionally brings it about. But intentional wrongdoing is not the only recognized form of wrongdoing. Agents may also be at fault for an action or outcome because they acted recklessly or negligently. A negligently produced outcome is not the product of a plan or desire, for agents who act negligently typically do not desire a harmful outcome. Nevertheless, even though negligent agents do not aim at wrongdoing, we sometimes hold them to blame for carelessly causing or not causing some state of affairs to obtain.¹

The objectionable quality of careless or dangerous action is not something new or peculiar to recent legal and moral theory. In Plato’s *Republic*, Socrates takes it as an accepted view that it would be unjust to return a borrowed weapon to a friend who has gone insane.

---

¹ Honoré (1999) argues that liability for negligence is not a function of culpability or fault, but of strict liability.
Presumably, Socrates’s justification is that returning the friend’s weapon would be needlessly risky. The Roman jurist Gaius writes: “A muleteer who lacks the skill to restrain his mules, so that they crush a slave, is ordinarily said to be liable for fault… since no one should undertake a task when he knows or should know that his infirmity will make its execution dangerous to others” (qtd in Honoré, 1999, p. 21). Thomas Aquinas dedicates an entire question of the Summa Theologiae to the particular sin of negligence, which he defines roughly as a lack of proper concern (II.2.q54). In article 3 of the same question, Aquinas goes on to argue that the sin of negligence can be every bit as serious as other types of sin.

Contemporary moral theorists agree that negligent conduct amounts to a failure to act with the appropriate level of care (Gardner, 2005, pp. 12-13; Raz, 2010, p. 9; Zimmerman, 1986, p. 200). However, they disagree on some specifics. Zimmerman (1986) argues that an action is negligent when, having acknowledged or adverted to a certain unreasonable risk in the past, an agent then forgets to avoid taking that risk (pp. 200-201). Other theorists argue that negligence does not require one to advert or acknowledge the risk (Gardner, 2017, p. 13). They argue that it is blameworthy to overlook a risk that one should have acknowledged. Anglo-American law tends to favor the latter, more objective interpretation of negligence doctrine: one is legally culpable for negligence if one fails to act with the level of care or diligence we would expect out of a reasonable person (Honoré, 1999, p. 86).²

² May (1992) at times favors a more subjective and at times a more objective standard of negligence. At one point, May argues that agents who “knowingly” engaging in risky or risk-raising behavior share some responsibility for any harmful outcomes that result (p. 45). At other times, May indicates that agents are to blame for being negligent if they fail to take “due care,” which sounds more like an objective standard of diligence (p. 43). When May goes on to discuss the negligent attitudes of racists, he seems unconcerned with whether or not those racists reflexively acknowledge the dangerousness of their attitudes, so I assume he ends up holding a more objective understanding of negligence.
Negligence often results from an omission to act. If I forget to signal before turning my car, I have failed to take due precaution because I omitted to switch on the turn signal and inform others of my intentions. But not all instances of negligence are omissions. If I drive to work exhausted, I am being negligent not because of something I forgot to do, but because my behavior is riskier than it ought to be. My “failure” to take due precaution is not a result of what I omitted to do, but rather with how I performed the action.

Philosophers and legal theorists also typically make a distinction between negligence and recklessness. While negligence results when one forgets or fails to advert to some risk in one’s actions, recklessness entails acknowledging that one’s plans entail substantial risks to others but acting on those plans anyway (Zimmerman, 1986, pp. 215-216). Recklessness involves a cognizant and willful disregard for the safety of others (Hart and Honoré, 1985, p. 214).

Negligence is typically a less severe form of wrongdoing than intentional wrongdoing. Intentionally breaking another’s arm deserves a more serious punitive response than carelessly causing a person’s arm to break. But this distinction is not absolute. Negligently bringing about a great deal of harm is more serious than intentionally causing some minor annoyance. To give an example, carelessly causing another person’s death is a more serious form of wrongdoing than intentionally hurting another’s feelings.

Anglo-American law allows that multiple negligent parties can contribute to the same wrongful outcome (van Dongen and Verdam, 2016). For instance, imagine that Noel and Joel each carelessly use a toxic fertilizer on their lawns and the fertilizer ends up seeping into a nearby lake, making the lake uninhabitable for wildlife and unusable for recreation. If Noel or Joel alone had used the fertilizer, then perhaps there would not have been enough seepage to contaminate the lake. However, because both carelessly used the fertilizer, their combined efforts
harmed the environment and any third parties who desire to use the pond for recreation. At times, English common law permitted aggrieved parties to sue any individuals at fault for negligence for full damages, while at other times courts divided liability to repair damages evenly among all culpable parties (van Dongen and Verdam, 2016, pp. 71-73). In recent years, it has become common for courts to assign responsibility for negligence according to the degree of contribution (van Dongen and Verdam, 2016, pp. 73-74). If one party’s negligence accounts for sixty percent of the resulting harm, then that party would be liable to remediate sixty percent of the damages.

**May’s Social Existentialism**

Drawing on Anglo-American negligence case law, Larry May (1992) develops a theory of how group members might share moral responsibility for unjust outcomes. He argues that negligent group members come to share in the intentional wrongdoing of their fellow negligent group members. If one negligent group member carries out wrongdoing, May argues that other similarly negligent agents share responsibility for that wrongdoing. More specifically, May argues that those who negligently adopt or maintain racist attitudes share responsibility for hate crimes perpetrated by other racists in their community. He believes this is the case even for racists whose attitudes did not contribute to any definite harm (p. 48).

May defends his linking of negligence and shared responsibility as follows. He argues that we are all responsible for the attitudes that we cultivate toward other human beings (pp. 23-24). For instance, if we practice and prioritize kindness and generosity, we ought to be praised for shaping our character in a way that is virtuous. Conversely, if we actively cultivate and

---

3 May is careful to say that those who share responsibility for the intentional actions of others ought not be liable to blame or feelings of guilt. Rather, May argues that group members who share harmful attitudes with those who commit wrongdoing should be liable to feelings of personal shame and moral taint (1992, p. 51).
reinforce hateful and bigoted attitudes, we ought to be criticized for shaping our characters in a way that could be harmful to others. May grants that many of our attitudes are not voluntarily chosen, but result from our upbringing. In those cases, while we may not bear responsibility for attitudes we did not willingly adopt or reinforce, May still argues that we bear responsibility for not taking steps to correct those harmful attitudes, whether through psychotherapy or habituation (p. 69). So May believes we bear moral responsibility for those attitudes we did not willingly adopt, as long as we could reasonably be expected to correct them.

May further argues that those who adopt or fail to correct a racist attitude raise the risk of harm to others. The more a person cultivates hateful attitudes, the more likely the person will be to commit a hate crime when given the opportunity. Therefore, those who adopt or fail to correct racist attitudes are no different from those who drive carelessly or create dangerous environmental conditions. As May puts it: “the person with racist attitudes is like someone who aims a gun at another person and pulls the trigger but, unbeknownst to him, there is no bullet in the chamber” (p. 49).

May argues that in a community of racists, it is mere moral luck that one person with racist attitudes carries out a hate crime instead of another, as both cultivate the same hateful attitudes. He continues:

The fact that the gun does not go off in his hands, but it does go off in the hands of the next person to pull the trigger, does not eliminate his share in the responsibility for the harm. Both people who act recklessly share responsibility not just for the risk but for the actual harm. (p. 49)

For May, there is little moral difference between the racist who happens to carry out a hate crime and the racist who lacks the opportunity. For that reason, he posits that if a person maintains harmful attitudes, that person shares responsibility for the wrongdoing committed by other
members of her community who share those attitudes, because she easily could have been the
perpetrator of the hate crime under different circumstances. Every member of a community of
racists therefore shares responsibility for any hate crime carried out by any other member of the
community. May has in mind here not only Klan members and violent misogynists: he insinuates
that faculty members and administrators who fail to take racism at a university seriously should
also share responsibility for hate crimes perpetrated on campus (pp. 36, 48).

May understands his theory of negligence and shared responsibility as following from the
legal ruling of *Sindell v. Abbott Laboratories*. In *Sindell*, the court had to figure out how to
distribute liability for damages caused by DES, a drug meant to prevent miscarriages (p. 49).
DES ended up causing cancer for some children of the mothers who took the drug. The court
could not determine whose dose caused the miscarriages, so it assigned contributory fault to all
manufacturers of DES based on their market share (*Sindell v. Abbott Laboratories*, 26 Cal.3d
588, 1980, §IV.8). Those with a larger market share had to pay higher damages and those with a
lower market share paid less. Because all manufacturers acted negligently in selling the drug, all
of them bore some liability to compensate families of the children who developed cancer, even
those manufacturers who may not have caused any harm at all.

May argues that all who negligently maintain racist attitudes share responsibility for
wrongdoing carried out by other racists in their community, just as all manufacturers of DES
shared responsibility for the resulting harm. Before we move on to a critical assessment of
May’s theory, it is worth saying something about May’s use of the DES case. The *Sindell* ruling
is an application of market share liability, one in which the court could not determine who
contributed what magnitude of the resulting harm. Instead, the court had to hypothesize about
which companies imposed the largest antecedent risk on pregnant women who took DES. They
somewhat crudely determined this by examining how many units of the drug each company sold (Sindell v. Abbott Laboratories, §IV.8).

The court looked to antecedent risk because of epistemic uncertainty about who actually caused the harm (Sindell v. Abbott Laboratories, §IV.8). But absent epistemic uncertainty, the courts would have assigned liability in proportion to each company’s actual contribution, not based on the antecedent risks they took (see van Dongen and Verdam, 2016, pp. 73-74). Shared liability based on antecedent risk therefore appears to be a sort of pragmatic compromise or second-best option for the application of comparative negligence by courts that face epistemic limitations. In the majority opinion of Sindell, Justice Mosk wrote that he assigned liability by market share because: “From a broader policy standpoint, defendants are better able to bear the cost of injury resulting from the manufacture of a defective product” (§IV.8). Mosk did not appeal to the demonstrable individual culpability of the manufacturers of DES, but to the social utility of shifting the burden of liability. The ruling of Sindell therefore does not on its own demonstrate that negligence entails shared legal liability or moral responsibility. If anything, the opposite is true: courts appeal to shared liability based on antecedent risk only in those cases when causation is unclear. The doctrine of market share liability is an abnormality born of necessity, not a foundational legal principle. For this reason, the doctrine on its own gives us little reason to accept May’s broader moral generalization that those with negligent attitudes share in the responsibility for wrongdoing of other community members who share those attitudes. May’s theory will have to stand on its own merits.

The Standard Liberal View

Far from being an elaboration of standard moral and legal norms, May’s understanding of the relationship between negligence and shared responsibility marks a departure from standard
practice. To see why, let’s reconsider the hypothetical example of the pool drowning. Imagine that you come home exhausted one night. As you crawl into bed, you remember that you left the gate to your backyard open and the pool cover off. You think about closing the gate and covering your pool, but you decide to go to sleep and do it in the morning. You awake to the sound of police sirens. Tragically, as you were sleeping a thirteen year old boy wandered into your backyard and drowned in your pool. At first, you are distraught, believing that your negligence led to the boy’s death. However, the police officer tells you that this is not the case. The boy and an older friend decided to go for a midnight swim in your pool. While in the pool, they got into an argument over a romantic partner and the friend intentionally drowned the boy. You are saddened by the whole affair, but secretly relieved that you are off the hook, so to speak.

The preceding example should make intuitive the standard liberal view of the normative relationship between negligence and voluntary conduct. While your negligence in leaving the pool accessible and uncovered may have contributed marginally to the boy’s drowning, the normative relevance of this contribution is cancelled or overridden by the voluntary conduct of the boy’s friend. You may still be to blame for carelessly leaving your pool uncovered, but you are not to blame for the boy’s death.

For an articulation of the standard liberal view of the relationship between negligence and voluntary conduct, we can turn to Hart and Honoré’s (1985) *Causation in the Law*. As the title of their seminal work implies, Hart and Honoré understand wrongdoing by and large as the product of a causation relationship. While demonstrating that a person’s conduct has a causal connection to a certain wrongful outcome is neither necessary nor sufficient to establish that the person bears responsibility for it, a causal connection remains the most integral way in which individuals might bear responsibility for a wrongful outcome (1985, pp. 66-67). Typically, when
we wish to discover who is to blame for some state of affairs, we first look to who caused or brought about that state of affairs through an intentional act.⁴

There are a number of ways that an individual might cause or bring about a state of affairs through acting or omitting to act. For one, an individual can intentionally produce the state of affairs through her own personal conduct. If Sam wishes to break Zeke’s arm, he could go over to Zeke’s house with a baseball bat and hit him forcefully in the arm. On any theory of action, Sam’s act falls under the description of intentionally breaking Zeke’s arm with a baseball bat. There is a clear and direct causal link between Samuel’s intentions, his bodily movements and the harm to Zeke (see Davidson, 1963).

Alternatively, Sam could cause Zeke to break his arm by unintentionally or carelessly raising the risk of harm to Zeke. For instance, Sam could fail to shovel the sidewalk in front of his house when it snows (May, 1992, p. 44). If Sam forgets to shovel the sidewalk in front of his house, and Zeke slips and breaks his arm while walking across the snowy sidewalk, then Sam is responsible for Zeke’s broken arm, not because he intended to break it, but because his negligence played a salient role in bringing it about.

Let us return to the example of the pool drowning. Suppose you negligently leave your pool cover off and your gate open, because you are tired and want to go to sleep. If a child wanders into your yard and drowns in your pool, then you may be partly to blame for the child’s death, because your negligence played a salient role in bringing it about. However, in the

---

⁴ Linking action to causation works most effectively when an agent brings about a state of affairs through direct intentional action. Things become trickier when we attempt to connect a state of affairs to an omission. For instance, if Homer Simpson allows the Springfield Nuclear Power Plant to melt down because he is sleeping on the job, he brings about an outcome through his inaction. Hart and Honoré (1985) argue that such omissions have causal efficacy, but others contest this point; see Dowe (2001) and McGrath (2005).
preceding example, an older boy intentionally drowned the child in your pool. While your negligence may have played a role in permitting access to the pool for the child and his friend, it did not play a salient role in the child’s death. The salient factor in the child’s death was his friend’s intervening intentional efforts at drowning him.

According to Hart and Honoré (1985), this is the standard model of how responsibility for negligence relates to responsibility for voluntary conduct. A voluntary wrongful act typically “negatives,” overrides or cancels the normative significance of the causal connection between negligence and a harmful outcome (pp. 213-219). To give another example inspired by Hart and Honoré, imagine that Ruth carelessly drives twenty miles per hour over the speed limit on the highway. Josh, seeing Ruth’s car coming in the distance, attempts to commit suicide by throwing himself in front of the speeding car. Josh is injured, but survives the impact. Because his injury results from his own voluntary conduct, Josh has no claim to restitution against Ruth. His intentional action cancels or overrides any normative significance between Ruth’s negligent speeding and Joshua’s resulting injury.

According to what I have called the standard liberal view, voluntary conduct typically overrides the normative relationship of negligence to a harmful outcome. A group of racists might all be to blame for cultivating and expressing hateful attitudes. But until any definite harm has occurred, the racists will be to blame for the irresponsibility of their attitudes and expressions of those attitudes only, not any wrongful outcomes carried out by other agents. Imagine that one of the racists intentionally carries out a hate crime. According to the standard liberal view, the racist’s intentional action ought to cancel the normative relationship between the other racists’ attitudes and the harmful outcome. By transitioning from careless risk-raising to intentional wrongdoing, the racist who carries out the attack comes to own the wrongful outcome. Like the
negligent pool owner who is not responsible for the child’s drowning, the other racists are let off the hook for the intentional violence carried out by their fellow racist, even if they are not off the hook for cultivating blameworthy attitudes, expressing those attitudes or encouraging others to adopt racist attitudes. May’s theory of distributing shared responsibility directly calls this cancelling or overriding effect into question. He argues that the voluntary conduct of racists does not typically cancel out the normative relationship between the negligence of other racists and the harmful outcome resulting from an intentional action. Rather, the other racists share responsibility for the hate crime perpetrated by one of their community members, even if they did not personally carry out any wrongdoing. According to May, they are not off the hook, but normatively connected to the wrongful outcome through their negligence.

It should be clear now that May’s understanding of the relationship of negligence and shared responsibility for intentional wrongdoing represents a departure from the standard liberal view. However, this is not necessarily a knock against it. It could be that we ought to amend our moral and legal practices to more resemble May’s proposals. However, I think the standard liberal view is preferable to May’s model, for a number of reasons. I proceed to articulate why in the next section.

**Four Objections to May’s Theory**

By removing the cancelling or overriding effect of intentional action, May hopes to address issues of racism and sexism. May argues that racists ought to share responsibility for the

---

5 If racists intentionally encourage others to carry out hate crimes, their behavior is no longer correctly described as negligent. Intentionally encouraging another to commit wrongdoing counts as a form of incitement, which is a blameworthy form of intentional conduct.
intentional actions carried out by other racists. In another work, he also argues that men with hateful chauvinist attitudes ought to share responsibility for rapes carried out by other men with hateful chauvinist attitudes (May and Strikwerda, 1994, p. 135). But I argue that May's rejection of the cancelling effect of intentional action leads to troubling results. For this reason, we ought to resist May's emendation and stick to the standard liberal view as articulated by Hart and Honoré (1985) coupled with an authorization view of distributive collective or shared responsibility.

First, removing the cancelling or overriding effect of intentional action greatly increases the number of ascriptions of moral responsibility. When a community member commits intentional wrongdoing, every other member of the community who shares the wrongdoer's attitudes shares responsibility for that wrongdoing. This means that racists bear responsibility for all hate crimes in their community and chauvinists bear responsibility for all rapes in their community. But, all people prone to speeding will also bear responsibility for the negligent car accidents in their community, aggressive people will bear responsibility for the assaults in their community and resentful people will bear responsibility for all crimes of revenge in their community. It begins to look like many people bear responsibility for a good deal of the wrongdoing committed in their communities.

Such a drastic increase of moral responsibility ascriptions is not in and of itself a bad thing. But it could lead to troubling consequences. Holding another responsible for an outcome is typically a serious charge. Not only does it sanction certain informal moral responses, but a charge of being at fault for an outcome may also ground legal liability to punishment or remedial obligations. People take attributions of responsibility seriously because those attributions typically entail serious social repercussions. However, it could very well be that once we take up
May’s suggestion, people will take charges of moral responsibility, less seriously. If, because of my insensitive attitudes and moderately risky driving, I am to blame for all the murders, car accidents, rapes and domestic violence in my community, I could justifiably conclude that being responsible for an outcome is not all that serious of a charge. To keep moral responsibility as a practically interesting and functionally useful concept, it may be helpful to limit the extent to which people become responsible for wrongful outcomes.

May attempts to mitigate this worry by positing that a person who shares responsibility for another’s wrongdoing does not deserve to be punished or blamed for that outcome, but only shamed for it (1992, p. 51). Even if May’s theory does entail that most people bear responsibility for much of the wrongdoing that goes on in their community, perhaps this will not trivialize the concept of moral responsibility, because they cannot be blamed for that wrongdoing, but only shamed. But there is reason to think that this qualification does not mitigate worries about the proliferation of responsibility ascriptions. There is empirical evidence that shaming individuals is as psychologically hurtful and traumatic as blaming them (see Oakberg, 2016). So it is not immediately clear that being liable to shame is demonstrably less serious than being liable to blame. In addition, increased shaming may lead to the same issues as increased blaming: if we subject people to shame for a healthy portion of the wrongdoing in their communities, they may justifiably conclude that being worthy of shame is not so serious. Indeed, the success of a political candidate as shameless as Donald Trump provides some indication that shaming may already have lost some of the psychological force it once had, especially across partisan lines (see Holmes, 2018).

But in addition to worries about the proliferation of moral responsibility, by removing the cancelling effect of intentional action, May effaces the clear distinction between negligence
and intentional action. In law and morals, intentionally bringing about a wrongful outcome is usually considered a more serious infraction than bringing about the same outcome by failing to take due precaution. The cancelling effect of intentional action gives intentional wrongdoing an overriding quality that establishes its increased severity. If we remove the cancelling effect of intentional action, it is no longer so clear why we should see any normative difference between the two. Intentional action and negligence would just be two similar ways that an agent might become responsible for an outcome.

We see something like this in May's theory of action. May places a strong focus on the decisions that agents make about how to shape or how not to shape their characters. Once a person's character is solidified, the actions that person carries out are just probabilistic determinations based on the content of that person's attitudes. If a person is fiery and aggressive, she will likely cause more harm. If a person is caring and controlled, she will likely bring about less harm.

Once we see intentional action as a probabilistic outcome largely determined by attitudes, it ceases to be a compelling normative category. According to May's theory of action, when we evaluate a person's conduct, what we really want to know is what sorts of attitudes a person cultivates. What the person actually does is merely an expression of those attitudes, so it is not so relevant to our evaluation. The focus of our moral evaluation is primarily the agent's negligence or diligence in forming her character.

Some may cheer this result, but it seems to impoverish our capacity to make nuanced determinations about wrongdoing. In both criminal and tort law, it matters a great deal whether an individual brings about an outcome through intentional action or negligence. Collapsing these two categories together would require some heavy revisions to our legal system, but also to our
informal moral practices, which still heavily focus on intentional wrongdoing. I think that a good deal more argumentation is necessary to demonstrate that such a drastic revision would be desirable or morally required.

A third concern with May’s account is how we ought to draw two theoretical lines. First, it is unclear how to properly distinguish between members of a racist’s community and persons outside of the community. It is not clear whether “community” means a circle of friends or one’s social groups, neighbors or citizens. Because racists can network with other racists online, we might even say that some racists are part of a global network of racists. It is unclear whether May believes that people share responsibility for the actions only of their close friends or if they share responsibility for the actions of all people in the world who cultivate similar attitudes. The word “community” could mean either and anything in between, depending on how we draw the line.

The other line-drawing issue is how to establish a minimum boundary for what gets considered a harmful or hateful attitude. The term “racism” could denote a large range of possible beliefs and attitudes (Garcia, 1996; 1997; Mills, 2003; Shelby, 2002). A racist could be a person who occasionally makes racially insensitive remarks, but it could also be a violent person with hateful beliefs or malicious attitudes. A racist might just be an individual who fails to do anything about systemic or structural racial inequalities. Unless we draw some lines, we are lead to the conclusion that every person with moderately insensitive views on race shares responsibility for the hate crimes carried out by malicious racists. This would be extreme, to say the least. Conversely, it may be the case that only those racists who have extremely malicious attitudes share responsibility for the hate crimes of other racists. But then this drastically narrows the scope of May’s theory. Given May’s argument that racist university administrators
share responsibility for hate crimes committed by others on campus, I doubt that May would welcome such a narrow interpretation of his theory.

I see no easy way to draw a line either to delimit the concept of community or decide what counts as a sufficiently harmful belief or attitude. To produce a workable theory, we will have to draw those lines somewhere. I take it that intuitions are going to clash about where to draw those lines, especially if we give people across the political spectrum a say. Defenders of May’s theory therefore must propose a compelling solution to both line-drawing problems to prevent May’s theory from sliding into either extremity on the one hand or triviality on the other.

One final objection to May’s distributive theory of collective or shared responsibility refers back to the normative standards of fairness from chapter one: proportionality and the moral separateness of agents. Recall that the principle of the moral separateness of agents stipulates that theories of responsibility ought not hold some individuals responsible for outcomes that have no connection to what they intentionally do. The principle of proportionality dictates that we ought to assign responsibility in proportion to an agent’s wrongdoing or contribution to wrongdoing. May’s theory violates both principles of fairness.

May’s theory of shared responsibility violates the moral separateness of agents, as it ascribes responsibility to individuals for actions and outcomes that have no demonstrable connection to what they intentionally do. Recall that May claims that racists are responsible for the hate crimes committed by other racists in their community, even if they have never expressed their hateful attitudes (1992, p. 48). According to May, individuals share responsibility for the actions of others by sharing their attitudes or failing to avoid sharing their attitudes. In either case, it is not because of what they actually do or plan to do. So by definition, May’s theory violates the separateness of agents, although I suspect May takes this to be a virtue of his theory.
But May’s theory also violates norms of proportionate blame (or, in May’s case, proportionate shame). As I noted in the first objection, May’s theory greatly proliferates the volume of moral responsibility that each of us bears. In addition to possibly trivializing the concept of moral responsibility, May’s theory also ascribes unfair magnitudes of moral responsibility. A somewhat careless person who occasionally makes insensitive remarks does not deserve to be shamed as if he had raped someone or committed a hate crime. As I noted earlier, May’s claim that individuals who share responsibility should only be shamed (instead of blamed or punished) does not necessarily solve the problem of disproportionate blame. It is still disproportionate to shame a person apt to make an occasional insensitive joke for all of the hate crimes that occur in his community.

For these reasons, we have good reason to avoid May’s theory of shared responsibility and stick with the standard liberal view of how negligence relates to intentional wrongdoing. The cancelling or overriding effect of intentional wrongdoing performs an important normative function: it prevents the rapid proliferation of responsibility ascriptions, keeps intentional wrongdoing normatively distinct from negligence and ensures that agents are held responsible primarily for conduct that they intentionally author or authorize. But some may argue that these costs are worth bearing because the standard liberal view is subject to serious objections. To show that this is not so, I now turn to some potential objections to my analysis.

**Objections to the Standard Liberal View**

In January 2011, Jared Lee Loughner shot liberal congresswoman Gabrielle Giffords in the parking lot of a Safeway. Liberal pundits quickly pointed out that congresswoman Giffords had been put under crosshairs on a map produced by Sarah Palin’s political action committee (Montopoli and Hendin, 2011). Palin’s map utilized crosshairs to show contested elections where
Republicans had hoped to pick up seats. Many of these pundits thought that Palin’s map was negligently inflammatory, and that she deserved some of the blame for the shooting. Giffords herself claimed in an interview: “We’re on Sarah Palin’s targeted list, but the thing is that the way that she has it depicted has the crosshairs of a gun sight over our district, and when people do that, they’ve got to realize there are consequences to that action” (qtd by Ye He Lee, 2017).

Many find such ascriptions of blame for negligent incitement intuitive. If your irresponsibility incites me to commit wrongdoing, perhaps you bear some of the blame for my wrongdoing. But if so, then the standard liberal view cannot be correct.\(^6\) Recall that the standard liberal view postulates that the normative connection between an agent’s negligence and some resulting harm is typically cancelled or overridden if the harm was the result of an intervening intentional action. This entails that, while one may be to blame for one’s negligence, one is rarely to blame for the intentional actions another carries out on account of that negligence.

There is good reason to think that when it comes to blaming negligent individuals for the intentional actions incited by their negligence, the standard liberal view is correct and contradictory intuitions are mistaken. First, charges of negligent incitement are rarely applied consistently. Around six and a half years after the Gabrielle Giffords shooting, a gunman shot conservative congressman Steven Scalise during a practice for the annual Congressional Baseball Game. The shooter, James Hodgkinson, had volunteered for the 2016 presidential campaign of left-wing populist Bernie Sanders. This time it was conservative pundits, who largely denied Palin’s responsibility for Gifford’s shooting, who called upon Bernie Sanders to take responsibility for the Scalise shooting, on account of Sanders’s frequent calls for a political

---

\(^6\) I thank Nathan Wood for raising this objection.
revolution (Kutner, 2017). Not surprisingly, the liberal pundits who previously assailed Sarah Palin disagreed.

There is therefore reason to think that charges of negligent incitement are often made in bad faith. It is hard to determine whether people genuinely think negligent incitement makes one responsible for the actions of another or whether they simply find it to be a convenient way to heap shame or blame on their opponents. Therefore, we have some reason to distrust people’s intuitions about the moral significance of negligent incitement, as they rarely feel such strong intuitions about their own conduct or the conduct of their in-group. It could be that upon reflection and ideal deliberation, many who purport to have these intuitions would concede that those who negligently incite others are to blame only for being irresponsible and not for any of the resulting intentional actions that others carry out due to their irresponsibility.

In addition to reasons to be skeptical about negligent incitement intuitions, the strict application of a doctrine like negligent incitement may condone various kinds of victim-blaming. Often, when one hears word of an outrage visited upon someone, a few people inevitably find it acceptable to blame the resulting harm on the victim’s purportedly careless conduct: if so and so did not want to be attacked, he ought not have walked home through a bad part of town. If thus and such did not want to be sexually assaulted, she ought not have worn such impractical or revealing clothing. And so on.

While victim-blaming proliferates on internet comment boards, most moral philosophers find the practice detestable, with good reason: blaming the victim is a profoundly illiberal shifting of responsibility from the agent who performs the evil deed to the victim who has already suffered harm. It both absolves the wrongdoer from bearing full responsibility and legitimates the victim’s suffering. But once we reject the standard liberal view of how negligence
relates to intentional wrongdoing, it is not clear that we can easily dismiss the justification of victim-blaming. If I can come to share responsibility\textsubscript{f} for the intentional actions of others through my negligence, then it seems plausible that a victim, through his negligence, can come to share responsibility\textsubscript{f} for the intentional actions of his attacker. At the least, given May’s theory of shared responsibility\textsubscript{f}, I know of no easy way to foreclose this possibility. Alternatively, the standard liberal view rules out victim-blaming in most cases. If intentional action typically negatives the normative connection between a victim’s negligence and the harm she suffers, then victims will usually not bear responsibility\textsubscript{f} for harms intentionally carried out by others. A victim may be to blame for being careless, but she is not to blame for her own suffering.

Third, if we accept the doctrine of negligent incitement and agree that typically one’s careless actions entail responsibility\textsubscript{f} for incited intentional actions performed by other agents, we still run into all the objections from section three. Rejecting the standard liberal view proliferates attributions of responsibility, blurs the line between negligence and intentional action, creates some line drawing issues and violates principles of fair blame (or shame) distribution. These costs need to be weighed against the potential benefits of attributions of negligent incitement. As I have noted, people across the political spectrum often disagree about justified applications of negligent incitement. It is therefore plausible that, even if a theory of negligent incitement is defensible, only a small subset of negligent incitement attributions will turn out to be appropriate. Therefore, it hardly seems worth giving up on the standard liberal view and the moral protections it affords for the sake of the occasional justified attribution of negligent incitement.

One may also object here that the whole practice of assigning responsibility\textsubscript{f} only to some negligent parties for certain outcomes is arbitrary and unfair (Nagle, 1979, 28-29). Imagine that
two individuals are both speeding down the highway at thirty miles above the speed limit. The
first individual ends up crashing into a passenger vehicle and killing the driver. The second
individual, upon seeing the first driver’s accident, slows down and drives the speed limit. In this
case, the fact that the first driver caused the accident and the second did not was pure luck. It
therefore seems arbitrary and unfair to hold the first driver responsible for the accident while
letting the other driver off the hook.

There may be something illiberal or unfair about assigning responsibility to negligent
parties for outcomes that result from mere luck. Proponents of May’s view could argue that,
because his theory impartially assigns responsibility for harmful outcomes to all who act
negligently, it is fairer than the standard liberal view, which assigns blame only to those who,
due to bad luck, happen to negligently cause a wrongful outcome to obtain. So much may be
true, but assigning responsibility to all negligent parties for a given outcome is not the only
impartial way to distribute responsibility. It would be just as fair to assign liability for harmful
outcomes to none of the negligent parties. Instead, we could just blame (or shame) them for the
extent of their negligence. May has given us no reason to accept his more punitive impartiality
over one that is more merciful.

Things are greatly complicated by May’s theory of action. Because May understands all
intentional wrongdoing as the product of negligently harboring certain attitudes, if we assign
responsibility in a way that is purportedly blind to moral luck, we either acquire responsibility
for all outcomes intentionally produced by others with similar attitudes or we acquire
responsibility for none of them. From a legal standpoint, this conclusion is unworkable: we
cannot reasonably hold all aggressive people criminally responsible for every murder in their
community. As long as we value the criminal justice system in anything resembling its current
form, neither is it acceptable to hold no one criminally responsible for the wrongful deaths they intentionally bring about.\textsuperscript{7}

While May’s proposal may mitigate the effects of moral luck, it also greatly increases the burden of liability born by individuals who cultivate negligent attitudes, which creates additional worries of disproportionate blame. I concede that, assuming May’s theory of action to be correct,\textsuperscript{8} May’s solution may mitigate the effects of moral luck more than the standard liberal view. However, the fairest and most proportional solution would be to blame people for the extent of their negligence only. But this would require us to deny May’s theory of shared responsibility, which seeks to connect wrongful outcomes to all who act negligently.

If May is incorrect, and intentional action is not merely the result of diligent or negligent character formation, then the standard liberal view at least mitigates the effects of moral luck by primarily blaming wrongful outcomes on those who intentionally bring them about. While all action involves an element of moral luck, a person suffers far more from moral bad luck when she is blamed for an outcome that results from her inadvertent carelessness than for an outcome that she plans for and intentionally brings about. Had things turned out differently, the inadvertent carelessness likely would not have led to the harm in question. However, if an agent forms a plan to carry out wrongdoing, even if things turn out quite differently, it is still plausible

\textsuperscript{7}Although we could plausibly hold individuals criminally responsible only for the extent of their negligent attitudes and criminal intentions. Given our cognitive limitations, it is not clear how this would work in practice. We typically learn about people’s negligent attitudes and criminal intentions through their negligent or criminal behavior. In addition, systems of moral or legal sanctions that punish thoughts, intentions and attitudes are often the subject of dystopian works of literature, which raises concerns about the general desirability of such a program. However, this line of thinking already governs some of our legal responses to negligent behavior. For intrinsically risky activities, such as driving a car, we already require individuals to buy insurance merely to engage in the activity. This partially shifts the burden of liability from those who unluckily cause accidents to the broader car driving population.

\textsuperscript{8}This is a rather controversial assumption.
that the agent will account for those differing circumstances and find a way to bring about the wrongful outcome.

In reality, every theory of responsibility likely entails some moral luck. Even given a theory of action like May’s, whether an individual recognizes and eagerly corrects character flaws at least partially depends on her genetic inheritance and early social environment, both things that she has no control over. The question is not so much how to remove moral luck, but how to minimize it while balancing other important moral considerations. I have argued that the standard liberal view at least mitigates some of the effects of moral luck while meeting other criteria of fairness. Even if May’s theory mitigates the effects of moral luck more than the standard liberal view, it still violates other norms of fairness. Given that we are likely stuck with some form of moral luck, the mitigation of moral luck that come with May’s theory does not seem to outweigh the other substantial normative and conceptual problems already highlighted in this chapter.

But some may still worry that sticking with the standard liberal view lets too many racists and sexists off the hook for their harmful attitudes. May originally introduces his theory of shared responsibility to address what he sees as a lack of accountability for racism at his own university (1992, p. 36). If this lack of accountability stems from moral norms in accord with the standard liberal view, then we may have a reason to modify the standard liberal view to address various forms of bigotry.

But it is not clear to me that May’s theory of shared responsibility really does a better job than the standard liberal view of addressing problems of racism and sexism. First, according to the standard liberal view, it is still acceptable to blame individuals on account of their
irresponsible attitudes, even if you cannot blame them for the intentional conduct of others. So racists and sexists are certainly not let off the hook for their negligent attitudes.

In addition, May’s theory of shared responsibility increases responsibility attributions for racists and sexists from the powerful classes, but it also increases responsibility attributions for many from traditionally oppressed classes as well. Individuals from traditionally oppressed classes also sometimes harbor negative attitudes, so they too may become responsible for the crimes committed by other members of their community. Just as May’s theory of shared responsibility proliferates responsibility attributions among the powerful, it proliferates responsibility attributions among the poor and powerless as well. This will provide the powerful with greater justification for shaming already oppressed people for conduct that they did not intentionally author or authorize, and I take it that this is a rather unwelcome result.

Some may worry that the liberal moral norms I defend protect the powerful against criticism, and this may be true. But they protect the poor and powerless from unfair treatment as well. By weakening these norms, we do open the powerful up to more criticism, but we also subject the marginalized to justified criticism as well. Because I see no non-ad hoc way of preventing this result, I think we have reason to stick to the standard liberal view and resist May’s theory of shared responsibility.

Another objection to the standard liberal view is that many do feel justified in taking legal action against negligent parties, even if there is an individual who demonstrably caused the harm through his or her free and intentional behavior. For example, the family of Jessica Ghawi, one of James Holmes’s victims from his 2012 movie theater massacre, sued four online ammunition dealers for negligently providing Holmes with bullets (Paul, 2014). Even though
Holmes intentionally carried out the shooting, Sandy and Lonnie Phillips felt strongly that the negligent ammunition dealers deserved some blame for the massacre.

In response, the standard liberal view only establishes that typically intentional action negatives or overrides the normative connection between one party’s negligence and the intentionally caused wrongful outcome. A proponent of the standard liberal view could concede that in rare cases, an intentional action may not always override the normative connection between another party’s negligence and a harmful outcome. The standard liberal view only establishes a firm presumption in favor of the cancelling effect of intentional action. Conversely, May’s theory assumes the reverse: that typically, one’s negligence does make one responsible for the actions of others.

It is worth noting that in most cases, Anglo-American courts have stuck to the standard liberal view. Hart and Honoré (1985) propose what I have called the standard liberal view as an explanation of how common law courts typically operate. In the case of Sandy and Lonnie Phillips’s lawsuit, not only did the court reject the lawsuit, the judge ordered Sandy and Lonnie Phillips to pay over $200,000 in legal fees (Ingold, 2015).

**Conclusion**

Thus ends the conceptual work of my project. My aim has been to demonstrate that an authorization theory of distributing shared or collective responsibility is fairest and most rationally defensible. I have defended an authorization account against prominent intention accounts from Anglo-American analytic philosophy. I criticized Iris Marion Young’s social connection theory of responsibility, which seeks to replace theories of backward-looking distributive collective responsibility with a forward-looking theory of shared obligations. Finally, I cast doubt on May’s proposal to weaken our traditional liberal moral norms to make
room for his theory of distributing shared responsibility. In chapter one, I claimed that an
authorization theory of distributing collective responsibility was fair and coherent. But now, it
appears that such a theory is preferable to its competitors as well.

If the preceding discussion has no meaningful impact on the way we live our lives, it will
have amounted to a mere mental exercise. But indeed I think it should change not only the way
we relate to others, but also how we design our institutions. In the final chapter, I demonstrate
that an authorization theory of distributing collective responsibility can go a long way to
clearing up some disagreements in international criminal law. In bringing criminal charges
against individuals for war crimes and crimes against humanity, international courts have
appealed to a number of doctrines of individual liability for collective crimes: command
responsibility, joint criminal enterprise and the International Criminal Court’s “control” standard.
Courts have so far not settled on a favored doctrine. In the following chapter, I argue that all
three doctrines are flawed in some way, and that an authorization theory of distributing collective
responsibility can help us to amend them so that they are fair and rationally defensible.
CHAPTER FIVE

INDIVIDUAL LIABILITY FOR COLLECTIVE CRIMES

James Holmes currently sits in a high-security prison in Allenwood, Pennsylvania. It is no mystery why he is there: Holmes personally shot and killed 12 people and injured countless others at a movie theater in Aurora, Colorado on July 20, 2012. Bracketing questions of Holmes’s mental competence, he clearly committed the murders and we have no trouble designating him a mass murderer. But we also designate individuals who have not personally killed anyone as mass murderers. Radovan Karadžić, Germain Katanga and Jean-Paul Akayesu all sit in jail as convicted mass murderers. Yet, no court established that these individuals had personally killed anyone; indeed no court even attempted to prove this. Rather, all three were convicted as perpetrators of collective crimes, ranging from genocide to crimes against humanity and war crimes.

Attributing criminal liability to individuals for collective crimes proves to be a challenging endeavor. Often, the political and military leaders we believe to be the perpetrators of such crimes have only an oblique or distant relationship to the atrocities in question. International courts have responded by developing a number of legal doctrines of perpetration, which seek to attribute collective rights violations to single individuals. In this chapter I examine three contemporary doctrines of perpetration in international criminal law: the doctrine of command responsibility, the doctrine of joint criminal enterprise employed by the ad hoc tribunals (ICTY and ICTR) and the “control” standard favored by the International Criminal
Court (ICC). Each doctrine has its virtues, but I argue that each is subject to some of the same conceptual and normative problems that plagued the distributive theories of collective responsibility addressed in the preceding chapters of this dissertation.

The doctrine of command responsibility actually encompasses a number of moral failings, some of which defensibly ground the imputation of criminal liability and some which, I will argue, do not. The doctrine of joint criminal enterprise suffers from similar weaknesses as Kutz’s (2000a) theory of participatory responsibility. As I argue in chapter two, Kutz produces no compelling rationale in defense of his theory (see also Atenasio, 2018). This normative deficiency also afflicts joint criminal enterprise. In addition, the doctrine of joint criminal enterprise heaps excessive magnitudes of criminal liability on marginal participants, a result that courts have so far mitigated through prosecutorial discretion. The ICC’s control standard does not quite hit the mark either, for it is difficult to define control in a way that is both conceptually coherent and effective enough to differentiate marginal participants from primary perpetrators.

These worries, among others, threaten the legitimacy of the international institutions of criminal law that employ these norms. For that reason, international jurists have good reason to modify these doctrines to ensure that they can be applied fairly, impartially and consistently.

To ensure that doctrines of perpetration can meet norms of fairness, I argue that individual perpetration of collective crimes ought to be a function primarily of command or authorization. Those who command or authorize others, whether tacitly or expressly, to violate international law become the perpetrators of those collective crimes. Any individuals who did not plausibly command or give their authorization to collective crimes cannot be called perpetrators, even if they may still be charged with other forms of complicitous, reckless or negligent conduct.
To frame my argument, I provide some background on the evolution of international criminal law, beginning with the establishment of the IMT charter. I then survey the development of command responsibility, joint criminal enterprise and the ICC’s control standard. I proceed to argue that we ought to orient doctrines of perpetration primarily toward those who command or authorize collective crimes and I conclude by making some suggestions as to how we might modify existing doctrines of perpetration to bring them into accord with the normative framework of distributing collective responsibility developed in chapters one through four of this dissertation.

From the IMT to the ICC

While norms of international law date back approximately 400 years, courts have only recently begun holding individuals criminally responsible for genocide, war crimes or crimes against humanity (Darcy 2007a, p. 189; Kelsen, 1943). After the First World War, the Allied powers recommended the establishment of an international tribunal to try the German Emperor (William II) and others for violating legal and moral norms (Darcy, 2007a, p. 191). The tribunal never materialized, so German officials were tried by the German Supreme Court in Leipzig (p. 191). The Supreme Court tried only 12 out of more than 800 original suspects and gave out relatively lenient sentences, none requiring more than two years in prison (p. 191). The trials were widely viewed as a farce outside of Germany (Tusa and Tusa, 1984, p. 19).

The end of the Second World War brought stronger demands for international criminal trials. The U.N. established a War Crimes Commission in October of 1943 to investigate Nazi crimes and suggest a potential legal response (Darcy, 2007a, p. 192). In June of 1945, the main Allied powers convened the International Conference on Military Trials in London. After working over the summer, they adopted the International Military Tribunal (IMT) Charter in
August of that year (Hale and Cline, 2014, p. 265). The IMT Charter claimed jurisdiction over essentially international crimes, such as crimes against peace, war crimes and crimes against humanity (Darcy, 2007a, p. 194).

Twenty-two high-ranking Nazi officials were put on trial at Nuremberg in November of 1945 (Tusa and Tusa, p. 504). While the IMT Charter permitted rather expansive charges of conspiracy and criminal organization, whereby anyone who participated in the affairs of the Nazi Party or Government would be responsible for the crimes of other participants, the Tribunal assigned criminal liability based on the extent of each individual’s personal contribution toward some criminal end (Darcy, 2007a, pp. 218-219). In total, the Tribunal sentenced twelve defendants to death, three to life in prison and four to prison sentences ranging from ten to twenty years (Tusa and Tusa, p. 504). Three defendants were acquitted.

The International Military Tribunal for the Far East (IMTFE), established to address crimes committed during the Second World War in Asia and the South Pacific, felt considerably more comfortable in applying expansive conspiracy and common plan doctrines. The court charged twenty three defendants with conspiracy to wage an illegal war of aggression (Darcy 2007a, p. 220). A few months before the establishment of the IMFTE charter, an ad hoc tribunal tried General Yamashita for failing to prevent his troops from committing atrocities in the South Pacific (p. 302). Yamashita was not charged with entering into a conspiracy or intentionally doing anything wrong. Rather, he was found guilty of having “failed to provide effective control of [his] troops as was required by the circumstances” and sentenced to death by hanging (UNWCC, 1948, p. 3). The precedent set at the Yamashita trial has come to be known as the doctrine of superior or command responsibility.
In 1993, the UN Security Council established the ICTY, an ad hoc tribunal meant to address atrocities committed in the former Yugoslavia (Hale and Cline, p. 267). Beginning with the Tadić ruling, the ICTY articulated and utilized a doctrine of criminal perpetration called “joint criminal enterprise.”¹ Joint criminal enterprise effectively criminalized making a contribution to a criminal plan. By making a contribution to a common criminal plan, one becomes at fault for the contributions made by all other members of the plan. In some of its most prominent rulings, from Milošević to Karadžić and Mladić, the ICTY decided to forego looking for individually perpetrated crimes, seeking only to link the defendant to one of four joint criminal enterprises carried out in the former Yugoslavia.²

The UN Security Council established the ICTR, another ad hoc tribunal, in 1995 to prosecute crimes that occurred during the Rwandan conflict. While prosecutors hesitated to rely on joint criminal enterprise at first, they eventually embraced the doctrine (Darcy, 2007a, pp. 235-236). The Appeals Chamber even amended several indictments to include charges of participation in a joint criminal enterprise (p. 236). In the trial of Aloys Simba, the ICTR based its charges of criminal liability only on the doctrine of joint criminal enterprise (Prosecutor v. Simba, para. 385).

In 1998, 120 nations ratified the Rome Statute, a multilateral treaty establishing an International Criminal Court (ICC) (Hale and Cline, p. 269).³ The ICC claims jurisdiction over those who commit war crimes, crimes against humanity, genocide and crimes of aggression in a

¹ Mark Osiel estimates that 64% to 81% of all ICTY indictments between 2001 and 2004 relied on the doctrine of joint criminal enterprise (2005, p. 794 n. 5).

² For example, see Trial Judgement Summary for Radovan Karadžić, available at: http://www.icty.org.

³ The United States is not currently a signatory of the Rome Statute.
The signatory’s territory (p. 269). The drafters of the Rome Statute relied on the German Criminal Code to establish two norms of perpetration: indirect perpetration and co-perpetration in article 25 of the charter (Jain, 2011, pp. 184-185). Indirect perpetration entails that an authoritative figure exercises control over the physical perpetrator of some crime (Hale and Cline, p. 270). Co-perpetration entails that a plurality of individuals, upon making a criminal agreement, together carry out a shared criminal aim (p. 270).

With the *Lubanga* conviction, ICC Jurisprudence expanded the doctrines of indirect perpetration and co-perpetration to develop a hybrid theory of perpetration: indirect co-perpetration (p. 270). Drawing on the German legal theorist Roxin, the Pre-Trail chamber posited that the two modes of perpetration from Article 25 of the charter could be combined, and that one could indirectly become a co-perpetrator by exercising control over a hierarchical network that collectively fulfills the material elements of a crime. However, some jurists remain skeptical: in a concurring opinion on the *Ngudjolo* verdict, Judge Van den Wyngaert argued that indirect co-perpetration was a novel doctrine of perpetration that did not follow from the two separate modes of perpetration from Article 25 (Ohlin et al., 2013, p. 734).

Thus we see primarily three doctrines of individual perpetration emerge from the jurisprudence of international criminal law: command responsibility, joint criminal enterprise and the ICC’s control standard. The ad hoc tribunals (ICTY and ICTR) greatly favored joint criminal enterprise, while the ICC has preferred to utilize and develop the control doctrine. The Rome Statute affirms command responsibility as a valid doctrine of perpetration in Articles 27 and 28. While joint criminal enterprise is not explicitly mentioned by the Rome Statute, neither has ICC jurisprudence denied its status as customary international law (Guliyeva, 2008, pp. 66-69).
Yamashita and Command Responsibility

The doctrine of command responsibility attributes crimes committed by a commander’s subordinates to the commander. For example, if a commander’s subordinates unjustly murder civilians, then the commander may be held liable for crimes against humanity, assuming certain conditions obtain. Typically, a charge of command responsibility requires the existence of a demonstrable and effective commander/subordinate relationship, a culpable omission on the commander’s part and the commander either knowing or having the means to know what his or her subordinates are up to (Ambos, 2009b, p. 130).

The Yamashita trial remains to this day the most famous (and controversial) application of command responsibility. General Yamashita was in charge of Japanese troops in the Philippines during World War II. During this time, Japanese troops gruesomely murdered and mistreated civilians, raped women and wantonly destroyed property (Bassiouni, 2013, p. 341). A U.S. Military Commission put Yamashita on trial for the actions of his troops, but prosecutors were unable to prove that Yamashita had ordered the misconduct or even that he had known about it (p. 342). Despite this, the Commission accused General Yamashita of having:

unlawfully disregarded and failed to discharge his duty as commander to control the operations of the members of his command, permitting them to commit brutal atrocities and other high crimes against the people of the United States and its allies and dependencies, particularly, the Philippines. (qtd in Bassiouni, p. 341)

The prosecution argued that Yamashita must have known or had the means to know about the actions of his troop and that this was sufficient to attribute their crimes to him (Bassiouni, p. 343).

---

4 To give two particularly egregious examples, one soldier tossed in the air and bayoneted an infant and a gang of twenty soldiers raped a girl and cut off her breasts (Bassiouni, p. 342).
The Commission found Yamashita guilty of failing to prevent the crimes of his subordinates and sentenced him to death (p. 342). Yamashita’s defense counsel filed a writ of certiorari with the U.S. Supreme Court, but to no avail: The Supreme Court found that Yamashita’s trial was fair and upheld the ruling (see in re Yamashita). Yamashita was hanged on February 23, 1946, and the precedent of holding commanders responsible for the actions of their subordinates established at his trial now often bears his name (Bassiouni, p. 345).

Partly in response to general uneasiness about the application of command responsibility to Yamashita, jurists codified the doctrine of command responsibility into international law in Protocol I of the Geneva Conventions (1977) (Bassiouni, p. 332). Article 86, paragraph 1 of the convention states that:

The High Contracting Parties and the Parties to the conflict shall repress grave breaches, and take measures necessary to suppress all other breaches, of the Conventions or of this Protocol which result from a failure to act when under a duty to do so.

The article places obligations on officials and superiors to monitor the behavior of their subordinates to ensure that they do not violate any of the articles of the treaty. Article 87 of the Protocol extends this obligation to military officials, requiring them to prevent their troops from violating international law. With the further establishment of institutions of international criminal law, the obligation to prevent the crimes of subordinates evolved into a substantial doctrine of criminal liability (Bassiouni, p. 332). Officials who failed to prevent their subordinates from carrying out war crimes or crimes against humanity stood criminally liable not merely for a breach of duty, but for the crimes themselves (p. 362).

The contemporary doctrine of command responsibility ranges over a number of moral failings. First, any superiors who order or command subordinates to violate international law become liable for those violations (p. 332). But the doctrine also transfers liability for crimes that
superiors did not command. In some cases, courts have used the doctrine of command responsibility to attribute crimes committed by subordinates to superiors that those superiors knew of and failed to prevent (p. 353). This is typically called the subjective standard. Courts more often enforce an objective standard: superiors are responsible for the crimes of their subordinates when they “should have” known that those crimes would occur (pp. 332-333). Jurists typically justify the objective standard by appeal to its capacity to deter: presumably, no one is in a better position to prevent soldiers from carrying out war crimes than their superiors, so it is reasonable to hold those superiors to an exacting standard.

Insofar as the doctrine of command responsibility transfers criminal liability to superiors for the crimes they order or command, it is clear and relatively uncontroversial. Killing another differs little from ordering a third party to kill another (with a reasonable belief that the third party will understand those orders as compulsory). For this reason, jurists typically refer to the responsibility one bears for ordering a crime as “direct” responsibility for that crime (Bassiouni, p. 332).

But legal theorists name the responsibility attributed to superiors who merely fail to prevent the crimes of their subordinates “indirect” (p. 332). Indirect responsibility derives from a number of possible omissions: when a commander fails to prevent unlawful conduct, fails to establish conditions that would prevent unlawful conduct, fails to investigate unlawful conduct or fails to prosecute or punish unlawful conduct (p. 332). In each case, although the commander does not commit or order the commission of the crime, the commander either allows it to happen or fails to take reasonable precautions to prevent it or deter future violations.

Between the subjective and objective standard for omissions, we have two possible grounds of indirect responsibility: the commander knew of a subordinate’s crime or the
commander should have known of the subordinate’s crime. The subjective, actual knowledge standard proves to be more exacting than the objective standard, for knowing that one’s subordinates are committing crimes and failing to do anything to prevent those crimes requires relatively good channels of communication and information. Many commanders who lack these means may be excused from culpability. But the objective standard is more subject to skeptical concerns, as it is significantly more difficult to determine when a commander “should have” known about the actions of his or her subordinates, as this determination requires us to advance certain epistemic, normative and modal suppositions about what commanders reasonably owe to society and what sorts of actions are physically possible for them.

The biggest advantage of command responsibility as a doctrine of perpetration is that, if successful, it holds commanders and high-level political leaders responsible while exempting marginal participants from bearing liability for collective wrongdoing. However, a number of practical concerns have made prosecutors hesitant to employ the doctrine. First, it is often difficult to determine who counts as a subordinate and who does not. General Radislav Kristić argued at his trial that Ratko Mladić had created an alternate chain of command that went around him (Bassiouni, p. 364). The ICTY Appeals Chamber acquitted General Tihomir Blaškić of crimes committed during the Lašva Valley campaign because it was not clear whether the police and paramilitary organizations who committed the murders recognized his authority (p. 365).

In addition, some regimes have become so bureaucratized that there is no longer a clear chain of command. Many South American juntas plausibly had no single person who could be isolated as “in charge” (Osiel, 2005b, p. 1779). In Argentina, appointments and retirements in the junta leadership were fully organized according to a schedule (p. 1780 n. 127). Even heads of state were not much more than cogs in the bureaucratic process. In such scenarios, it is difficult
to hold individual junta members responsible as having authority or control over any clearly defined set of subordinates.

Finally, many contemporary conflicts emerge organically through a number of distinct groups and organizations related only in an indirect sense. While individuals such as Jean-Paul Akayesu and Jean Kambanda had a hand in directing the Rwandan genocide, the conflict emerged organically in a number of distinct locations from a variety of disparate causes (Drumbl, 2005, p. 570). This problem persists with modern terrorist networks, which frequently lack clear centralization. Because the doctrine of command responsibility requires a demonstrable chain of authority and control, it struggles to assign liability for crimes committed as part of an organic or decentralized attack.

For these reasons, the doctrine of command responsibility has fallen out of favor in contemporary international criminal jurisprudence (Bassiouni, p. 384). Prosecutors did not want to spend substantial resources only to end up acquitting prominent military and political leaders. As a result, courts have begun to favor doctrines of blameworthy participation, such as joint criminal enterprise and the ICC’s control theory of perpetration. Yet, despite its diminished popularity, the doctrine of command responsibility remains a part of customary and positive international law. It was included in Article 7 of the ICTY Statute and remains operative in articles 27 and 28 of the Rome Statute.

**Tadić and Joint Criminal Enterprise**

The two ad hoc tribunals set up to deal with the conflicts in the former Yugoslavia and Rwanda (the ICTY and ICTR) have relied on a legal norm called “joint criminal enterprise” (JCE) to distribute liability to perpetrators of collective crimes. JCE is a form of common plan liability, sharing characteristics with conspiracy doctrine (Pinkerton liability) in U.S. criminal
law and joint enterprise in English law (Jain, 2011, p. 162). Whereas conspiracy doctrine in the U.S. criminalizes the formation of a plan to commit wrongdoing with others, JCE criminalizes the contribution to a common plan involving the commission of a crime (Hale and Cline, p. 268).^5

The ICTY Appeals Chamber explicitly outlined the doctrine of JCE in the Dusko Tadić ruling. Tadić, although charged with a handful of crimes, was initially found innocent of the killing of five men in Jaskici because there was no evidence he had participated in the murders (Ainley, 2014, p. 416). However, the Appeals Chamber overturned the ruling, arguing that Tadić was liable for the murders due to his participation in a JCE. The Appeals Chamber outlined three components of JCE doctrine, now known as JCE I, JCE II and JCE III. JCE I assigns collective liability for the consequences of a criminal plan to all who knowingly contribute to that criminal plan (Prosecutor v. Tadić, para. 196). JCE II stipulates that all who knowingly further a criminal system or enterprise (such as a concentration camp) are liable for any unjust harm that results from the system’s operation (Prosecutor v. Tadić, para. 202). JCE III establishes that in cases of JCE I or JCE II, one is not only responsible for harms aimed at by the criminal plan or system, one is responsible for any unjustified harms that are a “natural and foreseeable consequence” of the plan or system (Prosecutor v. Tadić, para. 204).

JCE I functions when a group of individuals establish a criminal plan specifically aimed at killing or violating the rights of an individual or plurality of individuals. There are no discernable limits as to how big the group might be or how long the plan might persist (Danner

---

^5 As Bassiouni (2013, p. 375) aptly notes, because many U.S. jurisdictions refuse to charge an individual with conspiracy without a demonstrable contribution to further the common plan, this distinction may not be so substantial after all.
and Martinez, 2005, p. 111). For example, many of the JCEs identified by the ICTY spanned a period of several years (p. 112). This creates some difficulties with isolating anything resembling a single shared criminal plan. Smaller conspiracies usually entail communication and coordination among all or most members of the plan, but this would be impossible if the JCE encompassed thousands of people spread out across an entire country.

Assuming these difficulties can be met, culpability according to JCE I requires three components: first, there must exist a common plan that aims at the violation of the rights of a victim or victims. Second, the defendant must participate voluntarily in one or more components of the common plan. Finally, the defendant must intend to perform an illegal action or assist others in performing an illegal action (Bassiouni, p. 378). Any individual who fulfills these requirements may be held liable for the consequences of the criminal plan (including those brought about by other participants).

JCE II finds its basis in prosecutions of individuals who participated in Nazi concentration camps (Danner and Martinez, p. 112). JCE II does not require a common plan or agreement aimed at any specific end, but only participation in an institutional system of mistreatment. In the Tadić ruling, the Appeals Chamber intended JCE II to cover death and torture camps, but in principle the doctrine could be used to cover other systemic rights violations as well (Bassiouni, p. 378). Culpability given JCE II requires that the following conditions be met: first, there must exist a system of repression or ill-treatment. Second, the defendant must actively participate in the system. Third, the accused must be suitably aware of the system’s operations. Finally, the defendant must intend to perpetuate or further the system (p. 378).
JCE III imports a rather controversial component from common law conspiracy doctrine. According to the U.S. precedent established in *Pinkerton v. United States*, any participants in a conspiracy become liable for the “reasonably foreseeable” consequences of the common plan (Ohlin, 2007, p. 152). JCE III establishes an identical stipulation: any participant in a JCE acquires liability for all the reasonably foreseeable consequences that result due to the plan. In the case of Tadić, although the courts could not demonstrate that he had murdered or ordered anyone’s death, the Appeals Chamber argued that death was an inevitable result of his plan to expel Muslims from the village of Jaskici (Danner and Martinez, p. 106). Because Tadić demonstrably took part in the operation to ethnically cleanse the Prijedor region of non-Serbs, the Appeals Chamber imputed criminal liability to Tadić for the deaths of five men who died while Tadić’s subordinates attempted to remove them from Jaskici.

JCE ends up being a rather expansive doctrine of perpetration. As Ohlin (2011) notes, the doctrine requires very little in the way of physical criminal acts, focusing instead on the individual’s criminal intent to participate in collective wrongdoing. Unlike the doctrine of command responsibility, JCE requires no clear command structure or commander/subordinate relationship. It merely requires intentional participation in some criminal plan (JCE I) or institution (JCE II). In addition, because of JCE III’s extended liability, prosecutors have been able to charge individuals with crimes that bear only a slight or oblique relationship to any identifiable plan or organizational structure.

The doctrine of JCE has been subject to rather harsh legal criticism. One prominent theorist jokingly called it “Just Convict Everyone,” as it permits prosecutors to go after marginal participants with a threat of severe liability (Badar, 2006, p. 302). To this point, prosecutors have generally chosen not to pursue charges against marginal participants, but that has largely been
the result of prosecutorial discretion (Osiel, 2005b, p. 1772). There was no legal barrier that prevented the ICTY from turning its focus from military and political leaders to lower level bureaucrats and flunkies. Some worry that giving prosecutors such wide discretion in deciding who to charge may create suspicions of illegitimacy (Danner and Martínez, p. 96). There is already a concern by some that international criminal courts amount to a form of victor’s justice, a worry that plagued the Nuremberg trials as well. If prosecutors may arbitrarily choose to prosecute some participants but not others, it may be harder to sell the court’s impartiality.

In addition, while the ICTY Appeals Chamber insisted that JCE was merely a codification of norms of customary international law, this claim is at best misleading and at worst demonstrably false. After an exhaustive study, Kai Ambos concluded that only JCE I and a limited form of JCE II existed as norms of customary international law (Ambos, 2009a). The Appeals Chamber’s justification of JCE relied heavily on previous murder prosecutions carried out by small groups of individuals in which every group member was present at the time of the murder (Bassiouni, p. 379). It is unclear whether such precedents are analogous to the statewide JCEs articulated by the ICTY. If the bulk of JCE doctrine is not customary law, but a novel creation of the ICTY Appeals Chamber, then ICTY convictions based on JCE violated the legal principle of *nullum crimen sine lege* (no crime without a law). If participation in a JCE was not established as a crime during the conflict in the former Yugoslavia, then the ICTY would have had no grounds for charging members of the Serbian military command with participating in a JCE.

As the mandates for both the ICTY and ICTR have come to an end, so has the prominence of JCE in international criminal jurisprudence, at least for the moment. While the International Criminal Court has not denied the status of joint criminal enterprise as customary
international law, the ICC’s Pre-Trial Chamber preferred to outline a different theory of perpetration based on indirect control over the crimes of others.

**The ICC and the Control Standard**

The International Criminal Court (ICC) has made use of a “control” standard of perpetration, drawing heavily from the doctrines of *Mittäterschaft* (co-perpetration) and *Organisationsherrschaft* (organizational domination) from the German Criminal Code (Jain, 2011, p. 167-171). The ICC’s control doctrine developed from jurists’ attempts to properly interpret and apply Article 25(3)(a) of the Rome Statute. The article assigns liability not only to perpetrators who directly carry out the material elements of a crime, but to any individual who “Commits such a crime, whether as an individual, jointly with another or through another person, regardless of whether that other person is criminally responsible.” Article 25(3)(a) establishes two ways in which a person might share responsibility for a collective crime with others: by jointly perpetrating the crime with them and by indirectly perpetrating the crime through them.

In the *Lubanga* ruling, the Pre-Trial Chamber utilized the notion of “control” to elaborate on the specifics of joint or co-perpetration. The Pre-Trial Chamber argued that only those who have control over the elements of the crime and are aware of this control may qualify as joint or co-perpetrators of collective crimes (*Prosecutor v. Thomas Lubanga Dyilo*, para. 341). In other words, joint perpetrators are those who are in a position to frustrate the commission of the crime, whether through action or inaction (Ohlin et al., 2013, p. 727). Other participants are innocent of the crime or mere accessories.

The ICC further developed the control doctrine in the *Katanga and Ngudjolo* decision. In *The Prosecutor v. Germain Katanga and Mathieu Ngudjolo Chui*, the Pre-Trial Chamber elaborated on what it means for an individual to perpetrate a crime through another. They argued
that indirect perpetration amounts to having control over the “wills” of those who carry out the material elements of the crime (para. 488(c)). While controlling another’s will typically amounts to coercing or authoritatively commanding another to act, the Pre-Trial Chamber also stipulated that one can exercise control over an entire organization that collectively fulfills the material elements of the crime (para. 500). This interpretation (somewhat controversially) combined indirect perpetration with co-perpetration to create a hybrid form of organizational liability.

Liability as an indirect co-perpetrator requires that the following elements be met: first, an individual must exercise control over an organization that violates international law (Jain, p. 185). As noted, to exercise control means that an individual has the power, through action or inaction, to frustrate the organization’s commission of the crime. Second, the organization must be suitably organized according to a hierarchical structure. There must be a leader or leaders who exercise authority over fungible subordinates (p. 185). The final requisite element is that the organization must carry out the crime through “automatic” compliance (p. 185). Subordinates must follow orders or be liable to replacement by others who will follow orders. When these criteria are met, it is as if those in authority utilize the organization as a tool to achieve their own criminal ends, and the courts hold them liable accordingly.

In many ways, the control standard strikes a balance between command responsibility and joint criminal enterprise. Like command responsibility, the control standard typically singles out superiors and those in charge. Unlike certain iterations of command responsibility, it does so by appeal to something the superiors do, namely exercising control over an organization that collectively violates international law. Like joint criminal enterprise, the control standard assigns liability due to participation. However, the control requirement exempts many marginal
participants and subordinates from liability, thereby allowing for less prosecutorial discretion than joint criminal enterprise.

But the control doctrine is not without its own worries. First, it is not entirely obvious that the concept of control is the best way to understand Article 25(3)(a). Because the article contains no explicit language of control, the most likely source of this interpretation comes from domestic control doctrines articulated in the German and Spanish criminal codes (Manacorda and Meloni, 2011, p. 170). Indeed, the Pre-Trial Chamber extensively quoted and appealed to the German legal theorist Claus Roxin in *Prosecutor v. Katanga* (p. 171). So while control is one way to interpret what it means to perpetrate a crime with or through others, it is not the only plausible interpretation of Article 25(3)(a).

In addition, it is not clear whether indirect and co-perpetration may justifiably be combined to create a form of organizational liability. Germain Katanga’s defense counsel objected that indirect perpetration and co-perpetration amounted to two distinct forms of liability, and that they did not entail a doctrine of indirect organizational liability (*Prosecutor v. Katanga*, para. 490). The Pre-Trial Chamber responded that it was appropriate to interpret “with another or through another person” as an inclusive disjunction that could be read to mean either one or both. However, in a concurring opinion in *Ngudjolo*’s acquittal, Judge Van den Wyngaert denied that the language of Article 25(a)(3) of the Rome Statute supported indirect co-perpetration (Ohlin et al., 2013, p. 734). Wyngaert proposed that the inclusive disjunction, while playing a role in formal logic, ought not be viewed as the obvious reading of “or” as it is
typically utilized in natural language (*Concurring opinion*, para 60 n. 76). According to Wyngaert, the ordinary language reading of “or” means one or the other, but not both.\(^6\)

Wyngaert further argued that modes of liability are only legally combinable if the individual elements of each mode are independently provable (para. 62). Because indirect (organizational) co-perpetration does not require proving that a superior acts with and through every member of the organization, one can demonstrate indirect organizational liability without demonstrating that the elements of either indirect or co-perpetration have been met. According to Wyngaert, this indicates that indirect co-perpetration is a novel doctrine of liability and not a simple fusion of two modes explicitly included in the Rome Statute.

In addition to these legal worries, the control doctrine raises some conceptual concerns as well. To begin with, the language of controlling may be unhelpful in understanding how humans relate to each other. Outside of slavery or outright domination, most organizational relationships involve a combination of willing assent and social pressure. It is unlikely that compliance is ever “automatic” in any organization the way that a machine’s output may be automatic. This problem is made worse in decentralized conflicts, where many individual groups and persons contribute to atrocities outside of each other’s control.

Further, assigning liability on account of control over an organization leaves us at a loss with how to deal with individuals who lose control over an organization. An officer or political leader who loses control over a handful of paramilitary groups who violate international law may

\(^6\) In the American common law tradition, Judges typically utilize a canon of construction named “*expressio unius est exclusio alterius* (the express mention of one thing excludes all others)\(^{,}\) which dictates that listed items on a statute ought to be interpreted as complete. Applied to Article 25(a)(3) of the Rome Statute, this would likely exclude indirect co-perpetration through an organization, as there is no express mention of perpetration “with and through” others. Had the drafters of the statute intended to include it, presumably they would have said so.
not deserve any less blame on that account. Those who think that leaders who negligently or recklessly lose control over an organization that violates international law ought to be held accountable as perpetrators will find the doctrine of command responsibility to give a more satisfying answer here.

While at times the control doctrine may assign too little liability to superiors, it also greatly expands liability in other cases. Some marginal participants may exercise a significant amount of control over collective wrongdoing. Paul Tibbets flew the Enola Gay as it dropped a nuclear weapon on Hiroshima. There is a sense in which Tibbets had far more control over what was going to happen than Truman for a segment of time. Paul Tibbets, through his actions, had the power to frustrate the plan at a time when Truman had significantly less control. But this would make Paul Tibbets co-responsible for the bombing of civilians in Hiroshima along with Truman and other U.S. officials, and this seems to elide some meaningful differences in the way in which they all participated in the bombing.

Thus concludes our summary of the major doctrines of perpetration and co-perpetration in international law. International courts have made extensive use of each doctrine, and all three can make a claim to the status of customary international law, even if jurists and legal theorists disagree over the specifics. In what follows I utilize the conceptual framework developed in chapters one through four of this dissertation to assess the normative integrity of each doctrine.

Justice or Collective Punishment?

The basic framework of assessment I apply is identical to the framework guiding this dissertation: doctrines should assess liability impartially and fairly while correctly labelling each perpetrator’s behavior and assessing a proportional magnitude of criminal liability. Doctrines that either assess sanctions for crimes that bear no clear relation to a defendant’s intentional
actions or that assess a disproportionate magnitude of blame are normatively questionable and ought to be subject to revision.

In chapter one of this dissertation, I argued that a person may justifiably be held responsible for the acts of others when that person authorizes them to perform those acts on the person’s behalf. This can occur in one of two ways. Either, the person expressly authorizes an agent to act by performing a conventionally recognized action or uttering a conventionally recognized statement in the proper social context, or the agent tacitly authorizes others to act by sufficiently participating in a coordinated endeavor with them. I proposed that the threshold for successful tacit authorization requires that an agent freely and intentionally join or enter into a common plan, have adequate knowledge of the group’s goals, aims and activities and make a substantial contribution toward the realization of the group’s common plan. An agent acquires responsibility for the actions she authorizes because authorization relationships, like promissory or consent relationships, are morally transformative (see Hurd, 1996 and Atenasio, 2018).

The morally transformative effect of authorization relationships helps to explain how individuals may acquire responsibility or criminal liability for the actions of others who act freely and voluntarily. But it is also the case that sometimes soldiers commit war crimes and crimes against humanity somewhat involuntarily. Some soldiers may violate international law because they have been commanded to do so by an authority figure, a command that comes with a plausible threat of death or serious harm for noncompliance.

In addition to authorization, individuals also bear liability for the crimes they order others to commit (assuming that they are not merely passing on orders under threat of death or serious bodily harm). As I noted earlier, we need not wade into discussions of collective responsibility to understand why commanding wrongdoing transfers liability. A commander who orders his
troops to commit war crimes or crimes against humanity (assuming those orders will be understood as compulsory) differs little from a person who uses any other instrument to violate the law. Killing civilians by launching reliable rockets at them differs little from killing civilians by compelling reliable troops to murder them.

This gives us two defensible grounds for distributing full liability to individuals for collective crimes: commanding and authorizing. Any individual who either commands or successfully authorizes others to violate international law ought to be held liable for those violations. Individuals who do not command or successfully authorize others to violate international law may be held liable as negligent or tried as an accessory, but they are not rightly designated as co-perpetrators.

While I have argued that commanding and authorizing serve as defensible grounds for distributing collective responsibility, I will not here claim the negative existential, namely, that no other such grounds exist. There may well be alternate grounds for transferring liability, but I have so far argued in chapters two through four of this dissertation that defenses of existing theories based on intentional participation, social connection or negligent attitudes come up short. It is incumbent upon proponents of alternate theories of distributing collective responsibility to produce a better rational justification. Until then, we may safely proceed with a skeptical stance toward theories of perpetration that assign (collective) criminal liability to those who do not command, authorize or personally commit violations of international law (see Atenasio, 2018).

We begin with the doctrine of command responsibility. As noted earlier, this doctrine turns out to encompass a number of distinct moral failures. First, command responsibility holds commanders criminally liable for the crimes of their subordinates that they order. Command
responsibility also holds commanders criminally liable for the crimes of their subordinates that they knew about but did nothing to stop (the subjective standard). Finally, command responsibility transfers liability to commanders for the crimes of their subordinates that they should have known about, with the added stipulation that they also should have done more to prevent those violations or deter future violations (the objective standard).  

Insofar as the doctrine of command responsibility holds superiors responsible for the actions of their subordinates that they command, it is mostly free of normative worries. This does not entail that every particular example of commanding war crimes or crimes against humanity is simple or straightforward. Commanders often give vague orders, and they may do so under any number of institutional or social pressures. While international criminal law has traditionally not excused commanders or soldiers for acting on superior orders, the ICTY has allowed such concerns to mitigate penalties at the sentencing stage (*ICTY Statute*, Article 7(4)). But when men as cruel as Hermann Göring or Ernst Kaltenbrunner enthusiastically sign their names to manifestly illegal orders, it is hard to see any objection in holding them liable for the crimes of their subordinates.

But we cannot appeal to the preceding justification if commanders know of the crimes of their subordinates but take no effort to prevent or deter them. If a commander does not command his subordinates to violate international law, then we cannot use command as a grounds for

---

7 There is a fourth moral failing sometimes associated with command responsibility: failing to punish subordinates after they have violated international law. I pass over analyzing this iteration of command responsibility for the following two reasons: first, it is not included in Article 27 or 28 of the Rome Statute, the current operative document of international criminal law. Second, while it was included in article 7 of the ICTY statute, ICTY judges uniformly agreed that failing to punish one’s subordinates amounted to dereliction of duty, not co-perpetration (Sepinwall, 2009, pp. 263-268). Failing to punish as a doctrine of transferring liability therefore has no claim to the status of positive or customary international law. Sepinwall (2009) offers some interesting arguments for why it should be included in international criminal law jurisprudence, but I do not have the space to address her concerns here.
transferring liability. What follows is that the commander either becomes a perpetrator through authorizing his subordinates or he is innocent or a mere accessory to their crimes. Justifying the subjective (knowledge) standard of indirect command responsibility requires us to show that a commander who knows of the crimes of his subordinates and voluntarily chooses to do nothing to prevent or deter them effectively authorizes those crimes.

Given the framework of authorization developed in chapter one, there is a strong case to be made that a commander who knowingly and voluntarily permits his subordinates to violate international law does tacitly authorize those crimes. Recall again the criteria for successful tacit authorization: group or common plan membership, accurate beliefs about the common plan and a substantial contribution. A commander who consciously permits his subordinates to violate international law easily fulfills the first two criteria, assuming that the commander recognizes his troops as under his authority and his subordinates recognize him as in command. By stipulation, we assume that the commander correctly believes that his subordinates plan to violate international law. Whether or not he successfully authorizes their crimes hangs on the extent to which his omission may count as a substantial contribution.

Recall in chapter one, I argued that a substantial contribution is a contribution that has causal efficacy, one that is integral to the realization of the plan and/or one that is issued from a position of influence or authority. A commander’s conscious failure to prevent or deter violations manifestly fulfills the influence proviso, but less clearly the other two. Whether or not a commander makes a significant causal contribution depends on whether or not conscious omissions have causal efficacy. On some theories of causality they do, on some they may not

---

8 For a precise definition and explication of these terms, see chapter one, pp. 49-51.
(see chapter four of this dissertation, p. 146 n. 4). It is also challenging to determine whether a commander’s conscious omission is integral to his subordinates’ aims to violate international law. To show that the commander’s omission was integral, we would have to show that such an omission made a difference and would have made a difference in a variety of similar scenarios.\(^9\)

But I think these challenges are not insurmountable. If a commander has a clear obligation to prevent the misconduct of his subordinates, and if he remains in regular contact with them and has a full understanding of what they intend to do, then it is reasonable for his subordinates to infer that they have their commander’s support. This makes it plausible to say that that the commander’s omission makes a difference and contributes significantly to their plans. There are more details to work out, but I have here sketched one argument in defense of the claim that commanders who knowingly permit their subordinates to violate international law tacitly authorize their crimes, and we can justifiably hold them at fault for those crimes accordingly.

It is less clear whether a commander who merely should have known about and prevented the crimes of his subordinates tacitly authorizes those crimes. Recall that it is already controversial to posit that a commander’s conscious omission counts as making a substantial contribution. But if so, it is even less clear that a careless or irresponsible omission amounts to intentionally making a substantial contribution toward some common plan or group activity. It is therefore less plausible that an irresponsible commander without awareness of his subordinates’ actions tacitly authorizes his subordinates to act on his behalf, for he neither intentionally makes

\(^9\) In other words, the fact that the commander’s contribution made a difference to the outcome of the collective action cannot be attributed to mere chance or bad luck.
a substantial contribution nor has adequate accurate beliefs about what his subordinates are up to.

If so, it does not seem right to attribute their crimes to him.

Assuming that General Yamashita really did lack the ability to communicate with his troops, then the U.S. Military Commission was not normatively justified in attributing the crimes of his subordinates to him. The underlying principle of such an attribution appears to be that commanders can, through their negligence, come to share in the crimes of their subordinates, assuming that they have a pre-existing duty to prevent their subordinates from violating international law. This principle is akin to May’s principle of shared responsibility, which stipulates that group members who share negligent attitudes also share responsibility for any misdeeds carried out by the individual group members.

I already gave some reasons why we should be skeptical of May’s theory in chapter four, and many of the same criticisms apply to attempts to attribute the crimes of subordinates to negligent commanders. To rehearse one important criticism: if Hart and Honoré (1985, pp. 213-219) are correct to say that intentional actions that aim at an untoward outcome typically override the normative significance of an agent’s negligent contribution to the same outcome, then it seems like the subordinates’ intentional violation of international law should override the normative significance of the commander’s negligence or negligent contribution. If the connection between the commander’s negligence and the ensuing atrocity is overridden or cancelled, then we can justifiably blame the commander only for his irresponsibility, not for the war crimes or crimes against humanity carried out by his subordinates.

A successful normative defense of the objective standard of command responsibility hangs on whether the standing duty to monitor one’s subordinates suffices to transfer liability to negligent commanders for the crimes of their subordinates. I have already argued that
irresponsibly failing to prevent the crimes of one’s subordinates does not amount to tacitly
authorizing those crimes. Neither does it amount to commanding those crimes. Proponents of the
objective standard must therefore develop an alternative normative grounds for such an
attribution. They will have to defend the principle that those who negligently fail to uphold their
duty to monitor the conduct of their subordinates become responsible for the wrongs their
subordinates commit.

Interestingly enough, a similar principle is operative in the Anglo-American common law
tradition: the doctrine of respondeat superior transfers liability for the wrongful acts of
employees to their employers. But it transfers civil liability, not criminal liability. So the
principle is not strictly analogous. Also in the common law tradition, the doctrine of felony
murder transfers criminal liability to any member of a conspiracy or common plan for any
murder committed during the commission of a felony, even those murders that fall far outside the
scope of the common plan. Presumably, the justification for transferring liability rests on the fact
that each member of the conspiracy has negligently or recklessly risked that the common plan
would result in wrongful death. But the doctrine of felony murder is a highly unusual legal norm,
and many philosophers and legal theorists attack it as indefensible (see Binder, 2008, pp. 966-
967).10 This comparison probably raises as many concerns as it hopes to address.

10 Binder (2008) himself is not a member of this majority, as he attempts a normative defense of the doctrine of
felony murder. However, in (2011), Binder makes it clear that he aims to defend only a limited application of the
doctrine whereby those who engage in particularly dangerous felonies acquire liability for any (foreseeable)
wrongful deaths that occur. His argument is that the criminal purpose of intentionally participating in a felony
aggravates a felon’s negligent homicide to murder. But then this understanding of felony murder is not strictly
analogous to the objective standard of command responsibility. On Binder’s reading, felony murder attributes
liability for murder to individuals who have already committed themselves to a depraved, immoral or
dangerous end. The objective standard of command responsibility raises the liability for a negligent contribution to liability for
intentional wrongdoing by appeal to the duty to monitor, not by appeal to any depraved motive on the commander’s part.
These considerations do suggest two defensible modifications for the objective standard of the doctrine of command responsibility, both which have been suggested by legal scholars. First, while commanding and authorizing subordinates to violate international law may transfer liability for the crimes themselves, failing to prevent or deter those crimes might be a different type of crime (see Sepinwall, 2009, pp. 263-268). The wrong in question would not be war crimes or genocide, but a failure to prevent war crimes or genocide. This could either amount to a charge of dereliction of duty, which would traditionally assign jurisdiction over such a crime to a military tribunal, or a form of criminal negligence. The disadvantage of taking the first option is that many people feel that court martialing a negligent commander insufficiently responds to the gravity of the commander’s omission (Sepinwall, 2009, pp. 292-298). The disadvantage of the second option is that the crime of reckless or negligent omission is not clearly outlined in the Rome Statute, and neither is it clearly a part of customary international law (Robinson, 2012, pp. 32-33). While neither option is perfect, both do correctly label the commander’s conduct without violating any norms of fairness.

Mark Osiel (2005a) presents another viable option for the objective standard of the doctrine of command responsibility. He suggests that the doctrine of command responsibility might be used to transfer mere civil liability to an entire officer corps that fails to prevent the commission of war crimes or crimes against humanity. An officer corps, through their negligence (or possibly through strict liability), would become liable to make economic restitution for the harm caused by their subordinates. This would make command responsibility more analogous to the doctrine of respondeat superior and mitigate some of the worries that any commander’s punishment is disproportional to his contribution. If the officer corps is held
financially liable to make restitution together, this also helps to avoid accusations that any one commander has been unfairly labeled as a genocidaire or mass murderer.

Having surveyed and assessed command responsibility, let us turn to joint criminal enterprise. Recall that the doctrine of joint criminal enterprise transfers liability for a collective crime to all who intentionally participate in and further either a common plan (JCE I) or institution (JCE II) that aims at wrongdoing. As noted earlier, this doctrine has been subject to severe criticism (see Badar, 2006). We can easily see why: as articulated by the ICTY’s Pre-Trial Chamber, joint criminal enterprise transfers liability for collective crimes to marginal participants in a common plan or criminal institution, so individuals are held liable for war crimes and crimes against humanity even if they have done very little to help bring about any specific unjust outcome. Charging a cook at a concentration camp with crimes against humanity both violates the separateness of agents and norms of proportionality. It transfers liability to the cook for actions that bear no clear relationship to her actions and it assesses an enormous magnitude of liability to the cook that far outstrips what she did in supporting a criminal institution.

Participating in a common criminal plan or furthering a criminal institution does not require that one command the other participants to violate international law. Neither, as I argue in chapter one, does one necessarily authorize others to further one’s ends by contributing to their plans or institution. Tacit authorization requires that, in addition to joining a common plan or group, one knowingly make a substantial contribution toward some aim or end. Marginal participants or those who contribute minimally toward a common plan or institution therefore do not tacitly authorize the other members to advance their aims. Many of them have no authority over, knowledge of or communication with the other participants. It is thus not reasonable for
other members of the institution or common plan to infer that they have the authorization of
other marginal participants and we cannot justifiably transfer liability to them for each other’s

Because joint criminal enterprise transfers criminal liability to those who neither
command nor authorize violations of international law, a defense of joint criminal enterprise will
require the articulation of an alternate normative ground of distributing collective responsibility.
Jens David Ohlin has attempted such a defense in (2011). He argues that, insofar as group
members jointly intend to commit a given collective crime, each member may be held
vicariously liable for the commission of the crime (2011, p. 742). In other words, if an individual
instantiates an individual intention to bring about a criminal aim with others, we can justifiably
charge that individual with advancing the ends of the criminal plan. If, in addition to
commanding and authorizing the crimes of others, intending that others commit wrongdoing
serves as an adequate normative ground for transferring liability for their crimes, then the
doctrine of joint criminal enterprise will be defensible, because it transfers liability largely on
account of criminal intent.

The astute reader will recognize the similarity of Ohlin’s defense of joint criminal
enterprise and Kutz’s (2000a) intention theory of distributing collective responsibility. Both
distribute liability by appeal to mental states instantiated by participants in collective wrongdoing
with minimal reference to what participants actually do. But for this reason, Ohlin’s defense of
joint criminal enterprise inherits many of the normative objections raised against Kutz’s theory
of distributing collective responsibility in chapter two. First of all, assessing fault by appeal to
intentions alone leads to disproportionate liability assessments. Ohlin and Kutz assign
responsibility not merely for what agents intend to personally do, but for what agents intend (or
desire) *that* others will do. Because any participant, no matter how marginal, can easily intend that genocide or crimes against humanity will occur, it is easy for that participant to bear liability for genocide or crimes against humanity, even if what they do is mostly innocuous. But if a cook or janitor can become a war criminal by intentionally providing minimal support to an unjust regime, we have effectively abandoned any attempts to accurately label an individual’s conduct and assess to her a fair magnitude of fault or liability.\(^\text{11}\)

In addition, proponents of intention-state theories of distributing collective responsibility\(^f\) have offered no compelling normative justification for their position (see chapter two and Atenasio, 2018). Commanding others to violate international law transfers liability for subordinates’ crimes to the commander because commanding others to act with a credible threat of serious harm or death is similar to using any other tool to commit wrongdoing. Authorizing others to commit crimes on one’s behalf transfers liability because authorization agreements, like promises or consent agreements, are morally transformative. But no similar compelling justification has been suggested by those who argue that intending collective crimes entails individual fault for those crimes. Defenders of intention theories of distributing collective responsibility\(^f\) either assume such a principle holds without offering a normative justification or offer weak and easily defeated justifications (see chapter two and Atenasio, 2018). So without a

---

\(^{11}\) I don’t mean to imply that desiring or intending that awful things occur is innocuous. We can justifiably blame people for such immoral desires and intentions. However, in addition to making a smaller contribution toward some criminal end, there is another important difference between marginal participants who desire that wrongdoing occur and those who make themselves an integral part of a conspiracy to commit war crimes or crimes against humanity. Marginal participants have not put themselves in a position to personally carry out war crimes or crimes against humanity, whereas many leaders and commanders have spent their lives acquiring the sort of power necessary to carry out their immoral intentions. This provides an additional justification for treating those commanders and leaders more harshly.
better defense of the principle that intending collective crimes entails personal responsibility, for those crimes, Ohlin’s normative defense of joint criminal enterprise fails.

The solution to bringing JCE I (common plan enterprise liability) and JCE II (institutional enterprise liability) in accord with basic principles of fairness is one that has been suggested by a handful of scholars and, interestingly enough, some ICTY Trial Chambers as well: both ought to require a substantial or significant contribution (Danner and Martinez, pp. 150-151). Recall that tacitly authorizing another to act on one’s behalf requires common plan membership, sufficient accurate beliefs about the common plan and a substantial contribution to further the common plan. Merely making a marginal contribution toward a common plan or criminal institution does not amount to tacitly authorizing the other members of the plan or institution to act on one’s behalf. However, if a person knowingly makes herself a member of a common plan or institution and coordinates with others to make a substantial contribution toward the common plan or institution, given the scope of the institution or common plan, then it is plausible for other members of the common plan or institution to infer that they have the person’s authorization to act on her behalf. We could therefore justifiably hold that person at fault for the actions others take to further the common plan or institution.

As long as JCE I and JCE II assign liability primarily by appeal to criminal intent, they assign liability disproportionately. Because a soldier may have the same criminal intent to contribute to war crimes or crimes against humanity as a military leader, JCE I and II struggle to make a meaningful distinction between the two. Prosecutors at the ICTY and ICTR generally shied away from prosecuting low level offenders, but to do so, they had to exercise a good deal of prosecutorial discretion. As mentioned earlier, this wide exercise of discretion has led some to question the tribunals’ legitimacy (Danner and Martinez, p. 143).
Whereas JCE I and JCE II might be normatively justified by requiring a substantial contribution to distribute liability, JCE III raises a different set of worries. Recall that JCE III stipulates that contributing members of a common criminal plan or criminal institution become liable for any reasonably foreseeable consequences of the common plan or institution’s activities. JCE III assigns liability for untoward consequences that participants should have expected would occur, given the nature of the common plan or criminal institution. JCE III therefore resembles the objective standard of command responsibility, which transfers liability to commanders for the actions of their subordinates that they should have deterred. JCE III transfers liability for collective wrongdoing on account of a participant negligently or irresponsibly participating in a criminal plan or institution.

I have already expressed my concerns with transferring co-responsibility, or full liability for outcomes or actions to agents who act merely negligently or irresponsibly. The previously rehearsed objections to the objective standard of command responsibility apply here as well. It is worth noting that proponents of JCE I and II typically do not endorse JCE III. Ohlin, despite his attempts to justify JCE I and II, insists that JCE III ought to be eliminated (Ohlin, 2009). It is difficult to find anyone who comes to the defense of JCE III, either from the perspective of law or morality, and this is further evidence of its questionable normative standing.

We now look to the least developed theory of perpetration: the International Criminal Court’s control doctrine of indirect co-perpetration. The control doctrine assigns liability to individuals who indirectly bring about a wrongful outcome through their control over an organization that collectively fulfills the material elements of an international crime. In principle, the control doctrine appears to advance the aims of international criminal law while remaining consistent with basic demands of fairness and justice. It effectively assigns liability to leaders
and powerful individuals who control and direct criminal organizations to violate international law. The difficult normative question is whether this is all the doctrine does. If it expands liability beyond the political and military leaders who direct organizations to commit wrongdoing it may be subject to charges of unfairness. In addition, the control doctrine may be liable to normative criticism if it fails to identify key or core perpetrators as perpetrators, effectively letting them off the hook for their participation in international crimes.

Insofar as the control doctrine of indirect co-perpetration assigns liability to those who intentionally direct an organization to commit wrongdoing, either through compulsion or agreement, the doctrine is not subject to any objections. As noted earlier, there is a firm normative basis for attributing liability to those who command or authorize others to commit wrongdoing on their behalf.\(^{12}\)

But there is another scenario suggested by the *Lubanga* Pre-Trial Chamber’s understanding of control. As noted earlier, the Pre-Trial Chamber interpreted joint control to require the ability, through action or inaction, to frustrate the commission of the crime (Ohlin, 2013, p. 727). When we scan the constituent parts of the control doctrine suggested by the Pre-Trial Chamber (control over an organization, fungible subordinates, automatic compliance), nowhere do we see a requirement that a perpetrator actively direct the organization to achieve a specific criminal end. It is enough that the perpetrator have the power to frustrate the commission of the organization’s collective crime.

If a leader is accurately informed about what her organization is up to, then, as I argued earlier, we can hold her liable as a perpetrator if the organization violates international law. A

\(^{12}\) An analogous argument can be made about those who deceive others into violating international law.
leader who has control and authority over an organization, correctly believes that the organization plans to commit war crimes or crimes against humanity, has the ability frustrate the organization’s crimes with little cost to herself, yet still does nothing to stop the organization tacitly authorizes the members of the organization to carry out the crimes on her behalf.

But if the leader of an organization lacks accurate beliefs about what her organization is up to, it becomes significantly more challenging to determine whether she is still in “control.” In one sense, the leader may still have the power to frustrate her organization’s actions, because she could, if she wanted to, transfer her soldiers or paramilitary troops away from where they plan to violate international law. But without sufficient information about what her organization is up to, she has no reason to exercise that control. This raises a similar concern to one we saw in our discussion of command responsibility. A leader could have control over an organization of fungible subordinates that would automatically comply with her orders, but fail to stop them from violating international law because she is poorly informed about her organization’s plans and activities. According to a control theory, the leader without sufficient information may still fulfill the requirements to be liable as a perpetrator. But because the poorly informed leader does not command or authorize her organization to act on her behalf, we require an alternate normative rationale for transferring liability to her, a rationale which, I have argued, is still forthcoming. So if the control doctrine does not require a leader to be adequately informed about her organization’s plans and activities, it is subject to the same objections that afflict the objective standard of command responsibility.

In addition, there might be some marginal or low-level participants in an international crime who exercise a tremendous amount of control over the success of an organization’s aims. Recall again the case of Paul Tibbetts, the pilot tasked with dropping a nuclear weapon on
Hiroshima. Tibbetts was an officer, but he did not personally devise the plan to bomb Hiroshima. Yet, insofar as he had the capacity to frustrate the bombing of Hiroshima by turning around and flying home, he did have control over the operation for the six and a half hours of the flight from Tinian to Hiroshima. His control was a function of the integral role he happened to play in the bombing. If so, it seems like the control doctrine would transfer liability to him for the bombing of Hiroshima. But given that Tibbetts was merely following someone else’s orders, it would seem somewhat disproportionate to attribute all the wrongful deaths at Hiroshima to him personally, solely because of the singularly powerful position he happened to occupy for a time.

Finally, we also have the problem of the leader who directs an organization to violate international law, but does so while exercising very poor control over the organization. A political leader might convince a paramilitary group to murder civilians, even if the political leader has no real authority over the group members or the capacity to enforce any behavior through automatic compliance. Perhaps the political leader just gets lucky because her requests align with the paramilitary group’s general inclination to harm and kill members of some out-group. But if the political leader lacks effective control, then the control doctrine fails to transfer liability to her for the paramilitary group’s crimes. Here I think many will find that the control doctrine does not go far enough. If the political leader authorizes the members of a paramilitary group to violate international law on her behalf, it does seem right to hold the political leader liable for that violation, on account of the criminal agreement. Whether or not the political leader exercises extensive control over the members of the organization seems somewhat beside the point.

While I think much can be said for the ICC’s control doctrine, it runs into two problems: first, it attributes severe liability to some marginal participants who happen, through no effort or
decision on their part, to have a significant capacity to frustrate the ends of a criminal plan.

Second, it fails to attribute liability to leaders who, despite exercising weak to no control over an organization, still manage to direct that organization to violate international law.

One way to remedy these issues would be to modify the control requirement. Recall that the control doctrine requires control over an organization, fungible subordinates and automatic compliance. Instead of control over an organization, we could stipulate that liability for the criminal actions of an organization requires that one either intentionally direct (through command or agreement) the organization to violate international law or, with adequate authority over an organization, intentionally omit to prevent the organization from violating international law. This would exclude those marginal participants who have no role in directing the aims of the offending organization from liability. This change would also assign liability to individuals who direct an organization to violate international law while exercising weak or no control over that organization, thereby addressing two of the primary concerns with the control theory.

Conclusion

As international criminal law continues to develop, we have seen three doctrines of perpetration emerge: command responsibility, joint criminal enterprise and the ICC’s control standard. Each doctrine captures something valuable about what it means to perpetrate a collective crime, but each is subject to a variety of normative or conceptual objections.

While command responsibility justifiably transfers liability for collective crimes to military and political leaders who command wrongdoing or consciously permit their subordinates to violate international law, it ought not transfer full liability to those commanders who negligently fail to prevent their subordinates from violating international law. Negligent
commanders should be held responsible for either criminal negligence or dereliction of duty, but not co-perpetration.

The doctrine of joint criminal enterprise too liberally assigns liability for collective crimes to all who intentionally participate in a common criminal plan (JCE I) or institution (JCE II) for any reasonably foreseeable consequences of the criminal plan (JCE III). If the doctrine is to survive as customary international law, it ought to require a substantial contribution and assign liability only for those consequences expressly aimed at by the plan or institution.

Finally, the control doctrine struggles to assign liability to those who direct an organization while lacking extensive control over its members, and it may occasionally transfer full liability for collective crimes to marginal participants. It would be better to understand indirect co-perpetration as a function of consciously directing an organization to violate international law (or consciously omitting to direct an organization under one’s authority to avoid violating international law) rather than as a function of control.

Some may worry that the modifications I suggest will make it harder to bring perpetrators to justice, and this is certainly correct. By bringing the aforementioned doctrines of perpetration into accord with norms of fairness and proportion, we thereby decrease their reach. But I conclude with two considerations in response. First, the aim of international criminal law, to borrow a phrase from Hannah Arendt, is “to render justice, and nothing else” (Arendt, 2006, p. 253). Securing convictions for international crimes is indeed important, but convictions cannot come at the cost of ensuring a fair and impartial trial. Otherwise, proponents rightly deride the whole process as an illegitimate form of victor’s justice. Second, criminal convictions are merely one component of international law and transitional justice. Post-conflict efforts may also include truth commissions, reparations, civil penalties and institutional and political reforms.
Many of these efforts might be as effective as (if not more effective than) criminal convictions at helping a nation in conflict move on and restore the rule of law. A decrease in criminal convictions need not thwart the ends of global justice.
BIBLIOGRAPHY


VITA

David Atenasio received a Bachelors of Music from the University of Massachusetts Lowell in 2004. He received a Masters of Arts from St. John’s College in 2011, a Masters of Arts from Miami University in 2013 and a Doctor of Philosophy from Loyola University Chicago in 2019. David is currently a Mellon/ACLS Dissertation Completion Fellow. His work has appeared in *Res Publica, Social Theory and Practice* and *Routledge Press.*