Title
Essays Concerning the Social Dimensions of Human Agency

Author
Story, Daniel Ervin

Publication Date
2020

Peer reviewed|Thesis/dissertation

University of California
Santa Barbara

Essays Concerning the Social Dimensions of Human Agency

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Philosophy

by

Daniel Ervin Story

Committee in charge:

Professor Kevin Falvey, Co-Chair

Professor Matthew Hanser, Co-Chair

Professor E. Sonny Elizondo

September 2020

The dissertation of Daniel Ervin Story is approved.

Sonny Elizondo

Matthew Hanser, Committee Co-Chair

Kevin Falvey, Committee Co-Chair

September 2020

ACKNOWLEDGEMENTS

There are many people who deserve thanks in connection with this dissertation. First, I would like to thank my committee members, Kevin Falvey, Matthew Hanser, and Sonny Elizondo. In addition to supporting me in the process of writing this dissertation, these three individuals have been invaluable teachers, mentors, and friends to me. Each of them has played an important role in shaping me into the philosopher that I am today. Although he was not on my committee, I would like to thank Dan Korman for similar reasons.

Many other individuals contributed to this dissertation in the form of helpful comments or discussions at one point or another. I hereby thank all such individuals. It is not possible for me to mention by name everyone who assisted me in this way, but I would like to specifically thank Arnel Blake Batoon, Saba Bazargan-Forward, Chris Britton, Sherri Conklin, Brendan de Kenessey, David King, Blake Kyler, Alex LeBrun, David Mokriski, Abe Roth, and Seyed Yarandi.

Additionally, I would like to thank the two other members of my graduate school cohort, Damien Barnes and Sam Zahn. In my experience, graduate school can be bizarre, dispiriting, tough, and tedious from time to time. But our friendships and shared activities are a significant part of what for me made graduate school also pleasurable, interesting, fulfilling, and good. I am very glad I was able to spend six years of my life with philosophers such as these.

Finally, I would like to thank my wife, Catelynn Kenner. Catelynn has been kinder, more understanding, and more supportive during this process than I deserve. Moreover, she has graciously talked through many of the ideas in this dissertation with me over and over again. I hope that I can benefit her in some of the distinctive ways that she has benefited me in our future life together.

VITA OF DANIEL ERVIN STORY
September 2020

EDUCATION

Doctor of Philosophy in Philosophy, University of California, Santa Barbara, 2020 (expected)
Candidate of Philosophy in Philosophy, University of California, Santa Barbara, 2017
Master of Arts in Philosophy, University of California, Santa Barbara, 2017
Bachelor of Arts in Philosophy, University of Kentucky, 2013
Associate in Arts, Bluegrass Community and Technical College, 2011

PROFESSIONAL EMPLOYMENT

2020-present: Lecturer, California Polytechnic State University, San Luis Obispo
2015-2020: Teaching Assistant/Associate, University of California, Santa Barbara
2015-2016: Grader, University of California, Santa Barbara

PUBLICATIONS

“Joint action without robust theory of mind.” forthcoming *Synthese*.
“Interpersonal Moral Luck and Normative Entanglement.” *Ergo*, vol. 6, no. 21, 2019, pp. 601-616.

FIELDS OF STUDY

Areas of Specialization: Ethics; Philosophy of Mind
Areas of Competence: Feminist Ethics; Metaphysics; Epistemology; Logic

ABSTRACT

Essays Concerning the Social Dimensions of Human Agency

by

Daniel Ervin Story

My dissertation concerns the social dimensions of human agency. I reject the common and often unstated presupposition that agents have clear boundaries separating one agent from another. In my view, it often becomes strained to think of interacting agents as functionally discrete spheres of intentional activity. This is because agents regularly act on one another's intentions and for one another's reasons. When your intentions and reasons guide and sustain my activities, there is an important sense in which some of my practical mental states and actions are attributable to you as well as to me, and there is no way to sharply distinguish what you are up to from what I am up to without distortion. My agency and yours have become intertwined.

Consider a simple example. Suppose my colleague phones me and asks me to find a document in our shared office. I have no idea where the document is, so my colleague actively directs my search. First, she tells me to check the filing cabinet. Then, she has me check the bookshelf. Finally, she tells me to check her desk, wherein I find the document. She then asks me to scan and email her the document, which I do after hanging up.

My colleague and I jointly searched for the document. But we played very different roles in that search. My colleague was deliberating and directing; I was acting at my colleague's direction and was guided by my colleague's intentions and goal. There is a sense in which my actions were an expression of not just my agency, but of my colleague's as well.

It is as if my colleague's sphere of intentional activity was interpersonally extended such that it overlapped with or twisted around my own intentional activity. We were agentially intertwined.

The first part of my dissertation is dedicated to defending these ideas. After surveying some of the leading accounts of joint/collective action in Chapter II, I argue in Chapters III and IV that directives are ubiquitous and integral to good social functioning in joint actions. Most contemporary philosophers who have written about directives have thought of them as tools for giving others reasons. But this approach fails to capture the distinctive ways that directives and interpersonal authority characteristically shape practical thought. I argue that directing another is a way of communicating one's intentions for them, and typically when one complies with a directive, one adopts that intention without independent deliberation about what to do, leading to the sort of agential overlap just described.

In Chapter V, I apply some of these ideas to a problem in developmental psychology. Psychologists hold that joint action is developmentally prior to robust theory of mind. Yet leading philosophical accounts of joint action presuppose that participants have robust theory of mind. I argue that even without a robust theory of mind young children can and often do share intentions and participate in joint action by adopting the communicated intentions of more competent partners who structure and manage interactions for them.

In Chapter VI, I turn to purely moral matters. I introduce *interpersonal moral luck*, which occurs whenever another's action, qua action, affects one's moral status in a way that is outside of one's capacity to control. I then argue that agents who are susceptible to interpersonal moral luck often for that reason enjoy special claims against those who are the source of that luck. I call this *normative entanglement*. I suggest that if my views about agency are correct, both phenomena are widespread in human life. This has important implications for our thinking about the nature of moral responsibility and the norms that govern agents in collective contexts.

TABLE OF CONTENTS

I. Introduction

This dissertation is about agency. In particular, it is about the social dimensions of human agency. It will come as no surprise that I think agency is one of the most interesting and important philosophical topics a person can study. Agency is a central, undeniable, and ever-present feature of human life. In fact, this probably understates things; perhaps the most fundamental fact about us is that we are agents. So, to study agency is to study *us*. It is to study human life and to ask what sort of thing we are. In my view, this is why studying agency is choiceworthy.

But why social agency? Well, we are not just doers in the sense that each of us does things and shapes the world in each's own image. *We* are also doers in that we do things *together*. We make plans together. We go on walks together. We build houses together. We eat together. We wage wars. We get married. We play games. We are insatiably social agents. Thus, to understand human agency, we must understand social agency. The overarching goal of my dissertation study has been to try to understand some of the most important aspects of human social agency. Consequently, this dissertation represents four distinct explorations into important aspects of human social agency.

The first essay ("Practical Intimacy and Collective Action") functioned as my qualifying paper at UCSB and is meant as a sort of introduction to some of the core philosophical literature on shared agency and human sociality. My interest in this chapter is understanding the nature of collective (or joint) action. I discuss two of the most popular philosophical approaches to collective action, and then I discuss some problems with both accounts. Finally, I describe and defend a thesis, originally due to philosopher Abraham Roth, which recurs in my dissertation. *Practical Intimacy* is the view that one person can act on another's communicated intention without deliberating independently about that intention. In this introductory chapter, I argue that we can put this thesis to good use in solving some of the difficulties that assail popular accounts of collective/joint action.

The second and third essays ("Directives, Practical Thought, and the Boundaries of Agency," parts 1 and 2) are in my opinion the most ambitious and exploratory of my dissertation. In these chapters, I argue that we cannot without distortion think of human agents as entirely discrete entities with clear, impermeable boundaries. I argue for this claim by first emphasizing the importance of directives in human social interactions, and then showing that when a director directs an addressee using a directive or similar device, this should be understood as one person expressing an intention for an addressee, which the addressee can take up and act upon directly (in line with the Practical Intimacy thesis detailed above). When an addressee acts on a director's intention, I think that the addressee essentially functions as an extension of the director's agency. As a result, the lines between what the addressee is up to and what the director is up to blur. My ultimate conclusion in these two chapters is that almost all joint actions involve a blending of agency in this way.

In the fourth essay of my dissertation ("Joint Action Without Robust Theory of Mind"), I develop an account of how very young children without a robust theory of mind can participate in joint action. There is something to be explained here because most leading accounts of joint action presuppose that participants have sophisticated mentalizing capacities. Ultimately, I argue that children can do this by taking up the communicated intentions with a joint content of more competent partners and acting upon those intentions.

In the last essay of my dissertation ("Interpersonal Moral Luck and Normative Entanglement"), I explore the connection between shared responsibility and moral luck. I argue that shared responsibility exposes agents to a distinctive type of moral luck, interpersonal moral luck, which occurs whenever the actions of others affect an agent's moral status in a way that is outside of that agent's control. One reason that interpersonal moral luck is distinctive is that it grounds entitlements to demand that others not negatively affect one's moral status. Thus, I argue that the possibility of interpersonal moral luck affects the normative landscape of joint interactions.

How do all these chapters fit together? The central recurring theme throughout my dissertation is the idea that a person can extend the boundaries of their agency by issuing intentions to others. In the first essay, I show how this idea can (and has) been used to account for collective action in general; in the second and third essays, I flesh out and defend the idea and explore its connection to directives; in the fourth essay, I apply it directly to the problem of making sense of joint action in children. The last essay might seem to stand out, and indeed it does. But if I am right in thinking that the boundaries between agents are fuzzy and porous, then shared responsibility will be a very common phenomenon and one's moral status will to sometimes—often—be in the hands of others. Thus, the fifth essay can be thought of as a study of some of the ethical implications of the findings that appear earlier in my dissertation. Of course, this dissertation does not by any means contain a complete picture or theory of human social agency. But I think one of its strengths is that it explores a novel idea about the boundaries of human agency from different and complementary angles.

II. Practical Intimacy and Collective Action

II.I. Abstract

In this chapter, I introduce some of the difficulties with giving an analysis of collective action. I first detail two popular strategies for analyzing collective action, which I call the *reductive view* and the *plural subject view*, and I describe some of the problems that plague each of these strategies. I then present a new strategy for making sense of collective action, due to Abraham Roth, which I call the *practical intimacy view*. I argue that the practical intimacy view combines the best elements of both the reductive and plural subject strategies, and after describing what I take to be the most salient difficulties facing it, I argue that they are manageable. I conclude that *practical intimacy* delivers a viable way of making sense of collective action.

II.II. Introduction

Imagine that you are at the pool watching two swimmers swimming in a single lap lane. The two swimmers (call them Jim and Kim) are swimming in accordance with proper lap-swimming etiquette: they swim to the far end of the pool on the right side of the lane, and they swim back on the left side. They pace themselves such that Jim is always about one length ahead of Kim; if Jim slows down, then Kim slows down as well. They avoid gratuitous splashing when passing one another. Jim and Kim are undoubtedly swimming *together* in one sense of the word. They are in the same lane at the same time. They have, in virtue of a sort of strategic responsiveness, solved a coordination problem (how to effectively share the lane) by adopting a certain convention and thereby reaching a coordination equilibrium.[1] But in another,

[1] Lewis (1967).

more robust sense of the word, it is difficult to tell just from looking whether Jim and Kim are swimming together.

If Jim suddenly jumped out of the pool and sprinted toward the locker room, Kim might stop and call after him, puzzled. This would be evidence that they were swimming together. Alternatively, Kim might go on swimming, completely unconcerned. In principle, Jim and Kim might be completely unaware of one another. The two might behave in accordance with the norms of lap-swimming etiquette entirely automatically. If this were so, Jim and Kim wouldn't be swimming together in the sense that interests us here. Kim might perform the exact same swimming routine if she were alone in the pool, if she were swimming with a stranger and unaware of it, or if she were working out *together* with a partner.

This example illustrates some of the philosophical problems with giving an analysis of collective action. Roughly, collective action occurs when two or more individuals perform an action *together*, in the robust sense gestured at above. Although collective action is a common feature of our social lives, it is difficult to say in what it consists and how to individuate it from the merely cotemporaneous and adjacent action of individuals. Collective action cannot be individuated simply in terms of the bodily behavior of individuals, because the same set of individual behaviors might in one instance constitute a collective action and in another instance not.[2]

One natural proposal, which I will adopt, is to suggest that collective action must be individuated in part by the intentions of the participants. Let the *Intention Thesis* be the idea that at least in paradigmatic, small scale cases of collective action, each individual participant

[2] For a more detailed argument for this individuation claim, see Searle (1990) pp.402-403 and Gilbert (1990).

in the collective action has an intention that somehow involves or implicates that joint action.[3] While the Intention Thesis is not unanimously accepted,[4] it is very plausible, both pre-theoretically and upon reflection. Intuitively, one can't participate in a collective action unaware or unwillingly. This isn't to say that collective action never involves coercion or manipulation, nor does it require that all collective action be premeditated rather than spontaneous.[5] But the Intention Thesis does require that participants in a collective action participate intentionally.[6] Broadly speaking, we can't know if the swimmers are swimming *together* unless we know something about their intentional attitudes.

What are the natures of the attitudes that satisfy the Intention Thesis, and what sorts of other conditions, if any, are necessary for collective action? In this chapter, I will examine three attempts to analyze the attitudes that characterize collective action. Roughly, the approach of these strategies is to develop a concept of a "shared" or "collective" intention which then can be used to explain collective action. According to the *reductive view*, participants in a collective action share an intention if they possess related individual intentions with a special content. On the other hand, the *plural subject view* holds that the participants collectively intend to φ iff they are "jointly committed" to together constitute a "pool of wills" dedicated to φing. I argue that there are serious difficulties with both views. As an alternative to these leading approaches, I proffer the *practical intimacy view* first developed by Abraham

[3] Roth (2010), Section 1.
[4] For detractors, *see* Miller (2001) whose analysis turns on the participants having a shared end, as well as Butterfill (2012) and Pacherie (2013), who develop alternative accounts that don't require an advanced theory of mind and so can be instantiated by very young children.
[5] c.f. Bratman (2014) pp.37-39, who distinguishes between shared *cooperative* activity and shared *intentional* activity, and Gilbert (1989) p.410, who allows for "a good deal of coercion" and lack of forethought by the members of a plural subject.
[6] This point calls to mind Aristotle's observation that actions motivated by coercion or under duress are still in an important way voluntary *NE* 3.1.

Roth, according to which two individuals can share an intention issued from one deliberative source. I argue that practical intimacy combines the best of both of the reductive and the plural subject view, while avoiding some of the most serious pitfalls of each.

II.III. The Reductive View

II.III.I. Understanding the Reductive View

The most natural place to begin is with the conceptual resources associated with ordinary, individual agency. If we could provide an analysis of collective action completely in terms of our individual agency and individual intentional attitudes, then we could dispel much of the mystery surrounding collective action. One of the most well-known approaches in this vein is due to Michael Bratman. Bratman argues that, due to their fecundity, the conceptual, metaphysical, and normative structures of his *individual* planning agency are continuous with the structures required to make sense of shared agency and collective action.[7] For Bratman, intentions are planning mental states that, in virtue of their association with certain practical norms, guide future conduct and coordinate activity over time.[8] Roughly, shared intentions are analyzed in terms of the intentions employed at the level of individual agency (and beliefs) with a distinctive collective content, and relations between those attitudes. I'll refer to this as the *reductive view*.

[7] Bratman (2014) pp.3-9. Bratman lays out his theory of individual planning theory of agency in Bratman (1987).
[8] Ibid. ch.3, 8.

Bratman's proposal has gained a good deal of nuance as he has developed it over the years.[9] But the core idea is that you and I share an intention to *J* (where *J* is our activity), and are thereby disposed to perform a collective action, if:

1. You and I each intend that we *J*.
2. You and I each intend that we *J* in accordance with and because of each of our intentions that we *J* and the meshing subplans of each of our intentions that we *J*.
3. (1) and (2) are common knowledge between us.[10]

Bratman argues that this structure of attitudes plays the functional role of guiding and coordinating members' actions toward the collective action. (1) and (2) describe high-level "strategic" intentions, in that they set the framework for lower-level "tactical" reasoning about the particular actions each will perform.[11] Since in (1) *each* of us intends that we *J*, and this involves not just your actions, but also mine, (1) entails that one can intend things other than one's own actions. (2) implies both that we each intend that our *J*ing involves the intentional action of all the participants and that each of our sub-intentions involved in executing our intentions that we *J* (e.g. intentions involving means) are compatible with one another. (2) is important because it excludes cases where we do something together only in a very attenuated, non-collective sense. If I tie you up against your will, throw you in the trunk, and drive you to New York City, then we have not performed a collective action (even if we were planning on going together at some point).[12]

[9] For Bratman's earlier formulations, see part two of (1999). For a later formulation, see *Shared Agency* (2014).
[10] This is Bratman's *Shared Intention Thesis* as presented in "Shared Intention" in (1999).
[11] Gold and Sugden (2007) p.136; the authors develop a complementary account according to which the key difference between collective and non-collective action is that the former originates in a mode of reasoning that takes the group as the primary unit of agency in "tactical" practical reasoning.
[12] Bratman (2014) pp.48-52.

As an account of collective actions, Bratman's proposal looks dangerously circular because our activity (*J*) is part of the content of our intentions in the analysis. We can't helpfully explain collective action in terms of a shared intention, and then explain shared intention in terms of collective action. However, Bratman argues that act-types involving multiple individuals are sometimes neutral as to their intentionality, and hence not necessarily collective act-types. Some act-types like *getting married* are necessarily intentional, but others, such as *sharing a lane at the pool*, are ambiguous.[13] The thought is that basic cases of collective action involve an intentional content that is intentionality-neutral, and these basic cases provide a basis for a "conceptual ratcheting" that supports shared intentionality with respect to conceptually less basic intentionality-loaded act types.[14]

Bratman's proposal has a certain degree of reductive elegance. First, his analysis does not appeal to any irreducibly collective attitudes or entities (such as group minds), and this lends it a degree of ontological parsimony and theoretical economy not found in some other accounts.[15] Moreover, his account explains certain normative features of collective action in an unmysterious and straightforward way. Because *I* have an intention that *we J* in a way that respects your intentional agency (clause (2)), I am rationally committed and constrained, via the norms associated with individual intentional agency, to act in ways that cohere with your

[13] The same is true for individual act-types. *Raising your leg* is neutral with respect to its intentionality, but *writing a novel* is not.
[14] Ibid. pp.43-48. Whether the "conceptual ratcheting" Bratman has in mind is a live possibility is the topic for an entire paper which I cannot broach in detail here.
[15] For instance, Bratman's account contrasts with Searle (1990) and Schmitt (2003). Searle holds a view according to which collective intentions are "primitive phenomena" that are unanalyzable in terms of individual intentions, and Schmitt argues that a collective action can only be performed by an agent that is not an individual. Schmitt's view is especially extreme in contrast to Bratman's in that Schmitt argues that his own view implies an eliminativism about collective action, because such agents are either rare (perhaps limited to large organizations) or entirely nonexistent.

intentions and our collective action. For instance, I cannot sensibly make plans that are inconsistent with *our J*ing, plus I am under rational pressure both to do my part in our *J*ing and to avoid subverting your intentional contribution.

II.III.II. Problems with the Reductive View

One problem is that Bratman's proposal seems to flaunt a highly plausible rational constraint on intending. We've already seen that Bratman denies that one can only intend one's own actions. Nevertheless, if I intend that *X*, then I must believe that my intention that *X* settles the question whether *X*. Call this the *settle condition*. The underlying idea is that I can only intend *X* if I believe I have the right sort of authority and control over *X*.[16] So while I can't intend that the sun rise tomorrow, I can intend that I be prepared for a meeting. The problem is that according to the reductive view, both you *and* I intend that we *J*. And it's not clear how *both* you and I can rightly see our respective intentions as settling the question of whether *we J*.

Bratman endorses the settle condition, and thinks that he can answer this challenge. To understand his reply, we first have to distinguish *onset* and *persistence* dependence. Intention A is *onset dependent* on intention B iff the onset of B triggers the onset of A in a way that involves the rational functioning of the relevant party/parties. For instance, my intention to go to the library on Tuesday is onset dependent on my intention to go to the library every day this week insofar as the former comes about as a result of the latter. In contrast, A is *persistence dependent* on B iff A persists only if B persists, as a result of the rational functioning of the relevant party/parties. For instance, my intention to run twice a week might be persistence

[16] A closely related observation is that I can't deliberate about something that I can't settle. Velleman (1997b) p.32.

dependent on my intention to get regular exercise. A and B are *persistence interdependent* iff A is persistent dependent on B and B is persistent dependent on A.[17]

Bratman thinks that your intention that we *J* and my intention that we *J* each settle whether we *J* in virtue of their persistence interdependence. The idea is that my intention settles whether we *J* in part *by way of its support of your intention*. Here's his argument. For the moment, assume that we both intend our *J*ing. Let it be true that if we both intend that we *J* at the appropriate time, then we will *J* by way of those intentions. If, in conditions of common knowledge, I were to change my mind and abandon my intention that we *J*, then I would not act in the ways required for our *J*ing and you would be rationally required to give up your intention that we *J*. The same is true, *mutatis mutandis*, for you and your intention that we *J*. Hence, our intentions are persistent interdependent – they each persist only if the other does. Because intentions are generally stable things, other things being equal, each of us will continue to intend that we *J* if—and only if—the other intends that we *J*. This is just to say that, other things being equal, we will *J* iff I continue to intend to *J* (and the same, *mutatis mutandis* for you). My intention will appropriately lead to our *J*ing 1) by way of my own actions, and 2) in virtue of the fact that your intention is persistent dependent on my intention; all this is true for your intention as well. If all of these features of our situation are common knowledge between us, then Bratman thinks you and I can both rightly recognize that our respective intentions settle whether we *J*.[18]

Bratman's solution is inadequate for two reasons. First, it's not clear how the intention that we *J* can get going in the first place. Bratman thinks that onset dependence is not an

[17] Bratman dismisses onset *inter*dependence as "a puzzling idea." Bratman (2014) p.68.
[18] Ibid. 64-67.

essential feature of shared intentions, and that onset dependence is always asymmetrical.[19] If our respective intentions are not onset dependent on one another, then it doesn't seem as if, in deliberating, either you or I can rightly see ourselves as settling whether we *J*. If your intention that we *J* is onset dependent on my intention that we *J*, then it's not clear how you can rightly see yourself as settling whether we *J* in *forming* the intention that we *J*. And vice-versa.[20]

There is a larger problem lurking here, however. A *settlement* is a determination of a practical question (via a rational process/processes), and an issue is *settled* iff there is a settlement with respect to that issue. The most natural way of understanding an intention's settling whether X is to think of that intention as *constituting* a settlement in favor of X. Typically, what it is for me to be settled on X is for it to be settled whether X *and* for one of my intentions to constitute the settlement whether X.

Persistence interdependence does not make it the case that my intention that we *J* constitutes a settlement. Your intention and my intention *together* constitute a settlement whether *J* (i.e. they jointly determine whether *J*), but they do not themselves constitute a settlement, any more than a tire constitutes a car. At best, they *partially* constitute a settlement. What persistence interdependence implies is that my intention persists iff it is settled whether *J*. But this is surely not enough to satisfy the settle condition. Consider a counterexample.

> Mischievous Brain Surgeon: Bob the brain surgeon implants a device into your brain that makes it the case that you are in Brain State B iff the president intends to give a

[19] Ibid. 68.

[20] I take it that Bratman's response to this would involve a sort of bootstrapping argument. Because I know that my intention, if formed, *will* settle the issue for the reasons detailed in the last paragraph, I can see myself at the time of deliberation as settling the question regardless of whether anyone's intention is onset dependent on anyone else's. But this won't work, for reasons detailed below.

> speech. You are in Brain State B iff you intend that the president gives a speech. You know all of this, and you know you are not the president.

Surely your intention that the president gives a speech does not meet the settle condition. Your intention persists iff it is settled that the president will give a speech, but it does not *constitute* such a settlement. So the settle condition requires more than just that my intention persists iff the issue is settled. Hence, my intention that we *J* does not meet the settle condition.

Bratman could respond by weakening the settle condition, such that I only have to believe that my intention *partially* settles the question at issue. However, no simple and sufficiently general notion of a "partial settlement" or "partially settling" seems forthcoming. It's not at all clear what it would mean to *partially* settle some issue. But even tabling that concern, Bratman seems to want to hold onto the stronger condition, and for good reason. Plausibly, my intention that Santa Barbara County vote Democrat partially constitutes a settlement that Santa Barbara County votes Democrat (if anything does), but intuitively my intention does not satisfy the settle condition.

A second problem is that it's not clear Bratman's account can adequately capture the complete practical-normative structure of collective action. On his account, participants in a collective action are rationally constrained to act in ways that cohere with the projects of the group and with participants' intentional agency. But in an important sense, Bratman's participants are normative islands in that they are only bound by the norms of individual intentional agency. If participants fail to do their part in a collective action, or if they undermine or subvert another participants' agency, then, *ceteris paribus*, they only let themselves down by failing to live up to the standards of prudence. According to Bratman, nothing about the nature of collective action makes it the case that participants are nonconditionally committed (or even normatively related) to *one another*.[21]

Of course, you may be morally obligated to act in certain ways if, say, failing to do so would sufficiently violate your partners' expectations, or if you have promised to do your part. But moral obligations are not necessarily present in collective action. Tabling morality, if I abandon our joint project, then at most you can point out that I am being inconsistent. But there's some reason for thinking that's not quite right: perhaps if we are doing something together and I fail to do my part, then you are in a *special* position to legitimately complain, because I have a *pro tanto* commitment to you (and you have a corresponding commitment to me) not to gratuitously subvert or abandon our project. Perhaps this is true even if you have no moral obligation to continue with our project.

Bratman denies this. But to really get to the force of this objection and to understand Bratman's reply to it, we need to contrast his account with one that makes these commitments central. I will return to Bratman's rejection of special commitments in below.

II.IV. The Plural Subject View

II.IV.I. Understanding the Plural Subject View

Let's return to Jim and Kim. The two strike up a conversation and, realizing that they live in the same apartment complex, decide to jog home together. The two leave the gym and everything goes well initially, but soon Jim starts to jog faster and it becomes more difficult for Kim to keep up. Finally, with Jim several yards ahead, Kim calls out, "Wait! Jim, slow down!"

[21] c.f. Gilbert (1990).

Jim and Kim are jogging together. But Jim is making it more and more difficult for the two of them to successfully perform that collective action. Jim is doing something strange, possibly illicit in speeding up. We might say that Kim is in a *special normative position* with respect to Jim (and vice versa). If, instead of Kim, a stranger called out to Jim and asked him to slow down, we would probably think the stranger officious. But when *Kim* does it, we make no such judgment. It seems as if Kim is *entitled* to rebuke Jim for his action, and it seems as if Jim has a *special obligation* toward Kim not to speed up (I'll often say Jim is "contralaterally committed" to Kim in this way[22]). The two are normatively conjoined in virtue of their collective action, and this is true even if nothing of any moral importance hangs in the balance. This normative entanglement is central to the *plural subject view* which I attribute to Margaret Gilbert and David Velleman. I'll first discuss Gilbert's version of the plural subject view, and then Velleman's emendation of it.

Against Bratman, Gilbert argues that shared intentions are not constituted by a concatenation of appropriately related individual intentions. Rather, to collectively intend with someone is to constitute a *plural subject* with her by being *jointly committed* with her. 'Joint commitment' is a technical term of Gilbert's, but she sees it as an articulation of a quotidian, commonsense notion. Joint commitments are "holistic" commitments *of* two or more people that are not reducible to complexes of individual commitments. A joint commitment is a commitment on the part of all to act in a certain way *as a single body*. When two individuals become bound to one another in such a way, they commit their own wills to be "part of a pool of wills which is dedicated, as one, to that goal."[23] Joint commitments expand participants'

[22] This terminology was introduced by Roth (2004) p.363 as a shorthand for the special obligations being gestured at here.
[23] Ibid. p.7.

practical perspectives by generating standalone reasons to further the ends of the group, licensing inferences from "We have goal G" to practical conclusions about what *I* will do as my part in *G* without reference to my own, individually-held goals, projects, or preferences.[24]

Gilbert argues that parties in a joint commitment enjoy rights against all other participants that demand conformity. These rights exist in virtue of the fact that a violation of a joint commitment is a violation of a commitment that is irreducibly *ours*. In one place, Gilbert says that if one fails to conform to a joint commitment, one disobeys oneself and one's collaborators.[25] With mutual claims come corresponding obligations on the part of each party toward the others to act as if one is part of a single spatially disconnected subject (or body) acting in the way implied by the joint commitment.[26] These rights and obligations are entirely derivative of and dependent upon the holistic joint commitment, and they cannot be rescinded by anything less than unanimous consent.[27] In the paradigm case, joint commitments require the implicit or explicit assent of all parties to come into existence, such that no one is committed unless everyone is so committed. In this way, joint commitments bind parties simultaneously and interdependently.

In an important paper, David Velleman develops Gilbert's view.[28] Velleman points out that according to Bratman's reductive view, participants in a collective action aren't sharing an intention in any literal sense.[29] He also complains that it is not clear how Gilbert's joint commitments result in anything more than a few people with well-coordinated commitments,

[24] Gilbert (1990) p.8.
[25] Gilbert (2000) Section 4.
[26] Gilbert (1989) p.411 characterizes these obligations as "associational" or "political" in nature.
[27] Gilbert (1990) p.8.
[28] Velleman (1997b).
[29] Ibid. p.31.

much less a plural *subject*. However, he contends Gilbert's proposal can yield an account of how two individuals can literally share an intention.

Velleman utilizes John Searle's theory of (individual) intentions. For Searle, intentions are intentional states that are causally self-referential, i.e. states that represent themselves as causing an action or movement. For instance, the content of a prior intention might look something like this:

A: {I perform the action of raising my arm by way of carrying out this intention}[30]

A's satisfaction conditions require that 1) the action of raising my arm be performed, and 2) A itself *cause* the action of raising my arm.[31] Whereas other intentional states such as perceptual experiences have as part of their satisfaction conditions certain causal relations between the states themselves and the remaining conditions of satisfaction (e.g. for it to be veridical, a visual experience of a table must be caused by a table), intentions are special in that they have a world-to-mind direction of fit.[32, 33]

Adopting and extending this idea, Velleman describes intentions as representations that resolve deliberative questions by representing themselves as causing a certain action or behavior. Velleman argues that Searle's analysis applies just as well to representational *acts* as it does to representational states, because the former can bear the appropriate content and causal roles.[34] For instance, some decisions made out loud constitute an intention. Saying "I'm going for a walk" in certain contexts is equivalent to saying "This statement is an effective

[30] I use the words in braces exclusively to represent the content of intentions.
[31] The same goes *mutatis mutandis* for intentions-in-action.
[32] c.f. Anscombe's (2000) mistaken shopper, §32.
[33] Searle (1983) ch.3 pp.79-98.
[34] Velleman tentatively extends Searle's analysis to the persisting conditions of representational acts that preserve the causal power of that act, e.g. the memory of a representational act (1997b) pp.41-42.

intention to take a walk."[35] With this conception of an intention in hand, Velleman uses Gilbert's view to provide an account of how multiple individuals can literally share an intention by participating in a public representation that meets Searle's criteria.

According to Gilbert, two individuals share an intention only if they first express their readiness to be jointly committed to a certain goal. We can imagine an exchange going something like this:

Jim: "I'll go running, if you will."

Kim: "Then I will."

These two expressions each express a readiness to be jointly committed to running, so long as the other is so committed (Kim's expression is elliptical). Gilbert holds that this is sufficient to establish a joint commitment, and hence for each to become bound to the plural subject's running.

Velleman interprets this exchange without utilizing Gilbert's technical notion of joint commitment, describing the utterances as representational acts meaning something like 'I hereby frame an effective intention to go running that's conditional on your intending likewise.'[36] Under this reading, each statement ascribes to itself a conditional causal power, namely that of causing the speaker to go running given that the other expresses the same. Velleman argues that because the two individual speech acts *together* make it the case that the antecedent of *each* conditional is satisfied, the two expressions together add up to a single public representation, in this case a combined speech act, "that causes [their] actions by representing itself as causing them—a single token intention that is literally shared between

[35] Ibid. 43-44.

[36] Gilbert ((2003) p.53) explicitly disavows the conditional commitment model proffered by Velleman.

[them].”[37] The idea is that Jim’s statement determines Jim’s behavior (and represents itself as determining Jim’s behavior) in conjunction with, and only in conjunction with, Kim’s statement (and the same goes, *mutatis mutandis*, for Kim’s statement). Hence, the two representational acts *together* add up to one intention that is literally shared between them, which represents itself as causing Kim and Jim’s collective action.[38] Because an intention shared in this way is a kind of joint decision, and because a decision just is a commitment to some course of action, participants who share an intention are committed, as one and by the effort of all, to some course of action. [39] Although Velleman is never explicit about this, I take it that I am committed to you in virtue of our shared intention because the *both* of us are directly responsible for that intention, and the intention is irreducibly ours in the same sense that Gilbert’s joint commitment is.[40]

The core of the plural subject view is the idea that a shared intention involves contralateral commitments on the part of all participants to pursue the goal in question. According to Gilbert’s variation, the source of these commitments is a joint commitment, whereas in Velleman’s variation the source is a literally shared representational act which constitutes a publically accessible intention.

[37] Velleman (1997b) p.47.
[38] Ibid. pp.44-50.
[39] Ibid. 48-49.
[40] Whether this is enough in every instance to ground a *right* against you in the way Gilbert has in mind is unclear, but this is not Velleman’s chief concern in (1997b).

II.IV.II. Problems with Velleman's Variation of the Plural Subject View

The logical structure of Velleman's conditional intention model is obscure. The basic idea is that participants each express conditional intentions, which then add up to a single representation that constitutes a shared intention. First consider the expression,

B: "I intend to go running, if you intend to go running."

B can be interpreted using wide or narrow scope:

C: I intend {to go running, if you intend to go running}

D: I intend {to go running}, if you intend {to go running}

I doubt that the narrow scope reading is the correct one. D is not itself an intention, but simply a descriptive statement which says something about the ways the world could be. D could be true even if I have absolutely no inclination to go running, or if I've never met you. This is at odds with how one would typically understand an expression like B. On the other hand, if we interpret B via the wide scope reading C, then I am rationally committed to ensuring the truth of the conditional. That means I am committed to going running given that you intend to do the same (as long as I have the intention). For this reason, I think the wide scope reading is the correct one.

Note that if the antecedent of a conditional intention goes unsatisfied, we can assume that the bearer of that intention is not committed (in virtue of that intention) to perform the action in the consequent.[41] If the antecedent in B goes unsatisfied, we can assume that I don't intend to go running. Similarly, from

E: I intend to {go running, if it doesn't rain}

[41] Hence, conditional commitments behave more like biconditionals than conditionals.

We can conclude that if it *does* rain, I am *not* committed to running.[42]

Velleman suggests that we understand Kim and Jim's statements of conditional commitments as something like this: "I'll go running, on the condition that you intend *likewise*."[43] We can now understand Jim and Kim's interaction as an expression of these two intentions:

F: Jim intends to {go running, if Kim intends likewise}

G: Kim intends to {go running, if Jim intends likewise}

It's not immediately clear how to cash out "intending likewise" in the antecedent. One way that won't work is to insist that the antecedent in F is satisfied iff:

H: Kim intends to {go running}

A nonconditional intention like H would not intertwine with Jim's intention in the right sorts of ways to constitute a *shared* intention in the way Velleman has in mind. Instead, Velleman argues that the antecedent of F is satisfied iff Kim possesses an intention resembling F. *Prima facie*, this idea is precise enough: Jim intends to {go running given that Kim has a conditional commitment *of this very sort*}, and likewise *mutatis mutandis* for Kim. The problem is that the self-referential bit 'of this very sort' may not be sufficiently determinate to ground the content of the commitment.[44] Velleman contends that two intentions can have similar content in virtue of their determinate or potentially indeterminate content. So, the intentions F and G are similar enough to satisfy one another's antecedents because 1) they share determinate content, i.e. a

---

[42] To be clear, the antecedent does not describe the *preconditions* that must obtain for the action in the consequent to be possible. Instead, the antecedent describes situations in which the agent is committed, in virtue of that conditional intention, to perform the action. *see* Ferrero's (2009) distinction between the *preconditions* and *restrictive conditions* of conditional intentions.
[43] Velleman (1997b) p.45.
[44] "Intends likewise" can't be cashed out in terms of explicitly determinate content because this leads to a regress problem: Jim intends to {go running if Kim intends {to go running if Jim intends to {go running if…}}}.

commitment to run, and 2) both intentions have potentially indeterminate content in virtue of vague self-reference.[45]

Abraham Roth has argued that this sort of conditional intention proposal is not viable. Roth's first worry is that the commitments purportedly generated by the shared intention are toothless. There may be ways in which participants in a collective action can legitimately unilaterally rescind their commitments to one another. For instance, an unexpected change in circumstance might justify unilateral abrogation (for more on this, see Section II.V.). But if the commitments are to be at all meaningful, then they must prohibit participants from gratuitously changing their minds. The problem is that it doesn't look as if Kim violates any of her commitments by suddenly and gratuitously abandoning her conditional intention, because in doing so she makes it the case that *Jim's* intention is no longer in force. Jim was only committed to running given that Kim had a conditional commitment like his, and Kim was only committed to running given that Jim had a conditional commitment like hers. But if Kim gratuitously abandons her intention, neither of these conditions are met, and neither party is committed to running. When this happens, the whole complex of intentions falls apart without any contravention of the commitments central to the plural subject view. In this way, Kim's failure is *self-effacing*.[46]

Roth's second worry is directed at the self-referential nature of the conditional intentions. Sometimes intentions and commitments can be unproblematically conditioned on something. I can quite intelligibly intend to go to the beach, on the condition that it doesn't rain. In doing so I really do settle something – I am committed to a certain course of action should the world turn out a certain way. But other conditionals are more problematic.

[45] Ibid p.45 n.26; Velleman (1997a) pp.50-52.
[46] Roth (2004) pp.376-378.

J: I intend to {go jogging if I intend to {go jogging}}.

The problem with this sort of intention is that in virtue of being conditioned on itself, it doesn't settle whether I will go jogging; in fact, it doesn't seem to settle anything at all.[47] Roth argues that it is an essential or constitutive feature of an intention that it settles *something*, and because J doesn't settle anything, it really isn't an intention. J is pointless.[48]

Roth thinks that F and G are problematic in a way similar to J. Look at it this way: if Jim already intends as in F and Kim knows this, then Kim knows that the antecedent of her conditional intention will be satisfied (and thus she will be commitment to running) iff she intends to {go running on the condition that Jim intends likewise}, because only then will her intention be similar enough to Jim's to make it the case that Jim's intention will satisfy the likewise clause. So her intention is effectively conditioned on itself, and she knows that it is effectively conditioned on itself. Roth's worry is that an intention that is effectively conditioned on itself doesn't really settle anything at all, and thus violates the practical logic of intending. For this reason, Roth thinks that in conditions of full knowledge Jim and Kim can't both form their respective intentions because they can't both see their intentions as settling anything, and so shared intentions can't get going in this way.[49]

If either of Roth's worries are on point, it spells trouble for Velleman's version of the plural subject view. Let's turn now to Gilbert's variation.

---

[47] In this case, if I do decide to go jogging (and thus satisfy the antecedent), *that* decision will settle whether I go jogging, not J.

[48] Ibid.

[49] Ibid. pp.376-380. If Jim or Kim's epistemic situation is impoverished in certain ways, then they might be able to form their respective intentions. But if the plural subject view is a true analysis of collective action, then it must be possible for participants to come to know all this and still perform a collective action.

II.IV.III. Problems with Gilbert's Variation of the Plural Subject View

It is difficult to say to what degree Gilbert's technical notion of a joint commitment is susceptible to the criticisms levied against Velleman in the previous section. Gilbert recognizes the problems mentioned above, and thinks that her view escapes such worries.[50] However, there are independent problems with Gilbert's notion of joint commitment and the concomitant obligations. To see this, let's first return to Michael Bratman.

Bratman denies that nonconditional obligations or rights are necessary or jointly sufficient for shared intention, and hence he claims they are not constitutive of collective action. He argues they are not necessary because a group of people could share an intention to *J* while explicitly remaining *uncommitted* to *J*ing, reserving the right to withdraw unilaterally. "I *intend* to sing a duet with you on Sunday, but I won't guarantee it."[51] And he argues that a joint commitment which generates obligations and rights is not sufficient for collective intention because a group of individuals can enter into a binding joint commitment while having no intention whatsoever to fulfill that commitment. Agreements bind even if they are insincere, but in such a case no collective intention can plausibly be said to exist.[52]

If Bratman is right to think that we can have a collective action without contralateral commitments, then there might be some collective actions wherein the participants are *not* in a special normative position with respect to other participants. At this point, intuitions may diverge: on the one hand, we could hold with Gilbert that contralateral commitments are

[50] Gilbert (2014b).
[51] Bratman (1999) pp.126-127.
[52] Bratman (2014) pp.113-118.

essential features of collective action, and hence that any activity not characterized by these commitments isn't a collective action at all. On the other hand, we could hold with the reductive view that while contralateral commitments may be typical of collective action, they are not essential to it.[53]

The discussion seems to have reached a stalemate. But there may be other, more pragmatic considerations that bear on the issue. Gilbert contends that a joint commitment cannot be unilaterally rescinded. She says,

> If I am subject to a joint commitment, I am subject to a commitment of a sort I can only rescind in conjunction with one or more other people. It would evidently be quite appropriate in that case to say I was "obligated" to those others and that they had "rights" over me with respect to the actions to which I was committed. These others are, after all, in a position to object to my failure to perform those actions in terms of the commitment in whose abrogation they know they have not participated. To the extent that the commitment stands, I owe them my conformity.[54]

Recall that these commitments do not rest on moral obligations. They do not depend on whether withdrawing would harm others or disappoint their expectations, nor do they require that I make a morally-binding promise. Concerning this feature of Gilbert's view, Christopher McMahon observes:

> The important question is whether there is a place in our practical lives for reasons that can require one to do one's part in a shared action even when the other parties have

[53] One could take an intermediate position. For example, Facundo Alonso (2016) argues that it is a constitutive feature of shared intentions that they induce others' reliance and thereby generate defeasible (moral) obligations. Because of this, he argues that these commitments are a definitive and non-accidental, but nonetheless non-essential, feature of collective action.
[54] Gilbert (2000) p.58.

invested nothing in anticipation of one's performance and would not find their pursuit of their goals impeded in any way by one's withdrawal.[55]

McMahon worries that the obligations inherent in Gilbert's joint commitments, which demand performance unless unanimously rescinded, are not likely to enrich our lives in any meaningful way, and may end up doing the opposite.[56] Gilbert is clear that the obligations associated with joint commitment do not solely determine what one all-things-considered ought to do. There might be overriding moral reason not to perform a collective action, for example. But she *is* clear that one cannot legitimately unilaterally withdraw from a joint commitment, and thus one cannot unilaterally leave the group or cancel one's obligations toward one's fellow participants.

I think that McMahon is on to something, and that the impossibility of withdrawal is a highly unattractive feature of Gilbert's view.[57] While joint commitments and the concomitant obligations are important, they aren't plausibly so sacrosanct as to preclude legitimate unilateral abrogation in every case. All else being equal (absent countervailing moral concerns), one can have practical reasons that make one's legitimate unilateral withdraw from a collective endeavor acceptable, e.g. if my interests change drastically, or if an unforeseen but

[55] McMahon (2005) p.300.

[56] McMahon suggests we understand joint commitments as generating *practically defeasible* considerations in favor of participation, whose reason-giving force cannot be rescinded except by unanimous abrogation but that can be overridden by other (personal) reasons that bear on participants' continued participation (ibid. pp.299-305). I think this suggestion might be acceptable to Gilbert, but only if we think of these reasons as *heavily protected* in Joseph Raz's sense. Protected reasons provide a first-order reason to φ while excluding from deliberation contrary reasons to not-φ. Protected reasons prevail over those reasons that they exclude, and a heavily protected reason excludes most or all potential reasons from consideration, Raz (2009) pp.17-18. My criticism of Gilbert directly below would apply even if she were to adopt McMahon's suggestion.

[57] Notice that this worry is exactly the opposite of the one levied against Velleman's version of the plural subject view in Section 3.2.

incompatible opportunity arises. This is compatible with owing other participants a preemptory warning or apology. Gilbert sees herself as developing an idea that has application within a wide swath of "the multifaceted domain of the philosophy of social phenomena," including the practical domain.[58] But I contend that unilateral withdraw from a joint commitment must be normatively possible if that notion is to be realistically integrated into one's practical reasoning as a source of all collective action. Even when we are limiting ourselves to the social domain, we can't reasonably expect one's *practical* commitments, once formed, to be utterly unresponsive to one's own practical agency. If this criticism is on point, then Gilbert's technical notion of a joint commitment won't work as the foundation for all collective action.[59]

Let's take stock of where we are now. We have seen a few of the most salient problems associated with the plural subject view and the reductive view. On the one hand, the reductive view is theoretically elegant in that it can explain shared intentions, and hence collective action, using only the resources of individual intentional agency, but it's not clear that the reductive view is compatible with the highly plausible settle condition (which Bratman himself endorses); moreover, the reductive view cannot explain why participants in a collective action are in a special normative position with respect to one another without appealing to extraneous considerations. On the other hand, the plural subject view implies that collective action involves a very rich normative structure, but the structure of the reciprocal conditional

[58] Gilbert (2014a) p.1.

[59] It is informative to consider that Gilbert attempts to explain promissory obligation in terms of joint commitments. For Gilbert, promises bind the promisee in the same way that collective intentions bind participants (Gilbert (2014c)). Note that my criticism here is different than McMahon's. McMahon suggests that we might amend Gilbert's view to make it acceptable while holding on to the idea that one can't unilaterally rescind a joint commitment. I think that even on McMahon's understanding, Gilbert's view is too strong.

commitments on which Velleman's version rests is obscure, and Gilbert's joint commitments seem to bind too tightly.

Bratman's reductive view and Gilbert's plural subject view are the dominant theories in the philosophical literature on collective action. Despite the fact that they each capture a good chunk of theoretical desiderata, the problems associated with each support a search for alternative strategies. In Section II.V, I outline a relatively new and underdeveloped strategy for making sense of some cases of collective action, due to Abraham Roth. According to *practical intimacy*, two individuals can share a single intention that is issued from one deliberative source. There's good reason to think that we often share intentions in this way, and that doing so generates contralateral commitments between sharers. I argue that this phenomenon can appear both against a background of an ongoing collective endeavor (as Roth usually describes it) and at the beginning stages of collective action, where it can explain how participants come to act *together*.

II.V. The Practical Intimacy View

II.V.I. Understanding the Practical Intimacy View

After their jog, Jim and Kim decide to go on a road trip together. On the day of the trip, Kim says to Jim: "Why don't you drive first, I have a headache." There are two ways to interpret what Kim is doing. On the one hand, Kim could be *suggesting* that Jim drive first, and Jim could take it as such. If Kim's statement is a suggestion, then she hasn't decided or settled anything for him. Rather, she has given Jim a reason in favor of driving first, which can then feature into his practical reasoning.

Alternatively, we might understand Kim as settling the question as to who will drive first, and thus *deciding for the two of them* who will drive first. Jim himself might see the matter as settled and, without weighing reasons for or against driving, simply do it. If Jim sees the issue as settled, then it wouldn't make any sense for him to deliberate about it, because one can't sincerely deliberate about an issue that one sees as settled (unless one reopens the issue by unsettling it).[60]

Consider another case. Imagine that you are walking down the street and up ahead you see a major car crash. You rush to the scene and someone tending to the wounded turns to you and yells, "Call 9-1-1!" At this point, there are several things that could happen. You could decide for yourself what to do (perhaps quickly and unreflectively), taking into consideration your *prima facie* judgment of the commander's reliability, your beliefs about these sorts of situations, and your desires. Or, you might just do it. The commander's command might *in itself* settle what you will do.[61]

These examples illustrate the central idea behind the *practical intimacy view*, which I attribute to Abraham Roth. According to practical intimacy, one person (the issuer) can issue an intention to another (the addressee) that includes as its content actions of the addressee, and the addressee can act directly on that intention without "reissuing," or deliberating about, that intention.[62] More generally, the idea is that in certain circumstances you can settle practical issues for me without any decision on my part. In this way, the practical intimacy view holds that you and I can literally share an intention that, say, {we go swimming} if you form an intention for the both of us and communicate it to me to act on directly. The thought is that

[60] Roth (2003).
[61] Roth (2004) p.387-388.
[62] Ibid. p.383.

when you decide for the both of us that we will go swimming, both you and I are settled on and committed to swimming, and an intention with this content can be attributed to the both of us despite the fact that the normative and causal *source* of that settling are the reasons and deliberations in your own head. Your intention for me features into my practical thought just as my own intentions do.[63] So the practical intimacy view attempts to solve the problem of how multiple individuals can share an intention by holding that two (or more) individuals can have an intention that originates in a single agent's deliberations. [64]

Roth argues that practical intimacy can satisfy our intuitions regarding the contralateral commitments constitutive of collective action. The basic idea is that when I take up your intention, I treat your intention as if it is my own, and my commitment to you consists in the fact that I am committed, in virtue of the norms of practical rationality, to act and intend in ways that cohere with your intention, just as if it were my own. Similarly, in issuing the intention to me you are committed to me to act in ways that cohere with your intention, and if you fail to do this, then you let me down.

These contralateral commitments depend on your *practical authority* to settle what it is that I will do, and my corresponding *entitlement* to your settlement as a practical conclusion.[65] In the intrapersonal case, I have the practical authority to settle for my future self what my future self will do, and I have a corresponding entitlement to my past self's decisions as practical conclusions (i.e. I can legitimately act on those practical conclusions without

---

[63] Roth (2014a) p.642.

[64] Roth argues that the content of the intention is *preserved* from the issuer to the addressee, just as the content of an intention formed in the past is preserved through memory. Roth draws on Burge's (1993/2013) discussion of how content can be preserved in testimony and memory. *see* Roth (2004) n.54, 56 pp.407-408**.**

[65] Roth thinks that I *cannot* settle an issue for you by issuing an intention in the way outlined by practical intimacy unless I have the relevant practical authority, and so the former always entails the latter (2004) p.390.

deliberating again). Likewise, sometimes *you* have practical authority to settle for me what I will do, and sometimes I have an entitlement to your settlement as a practical conclusion. The idea is that in certain circumstances you can legitimately issue an intention to me and I can take it up without deliberation, and in doing so I thereby become settled on some course of action. In such a case I have an intention that has originated in *your* practical authority, and in this sense it is not my own. Your intention plays the same normative role for me as my own intention would, but because it is bound up with your practical authority instead of mine, in failing to satisfy the practical norms associated with your intention I am letting you down by flaunting your practical authority. Conversely, when you issue an intention to me, you commit yourself not to act in ways that would undermine that intention, because if you did you would undermine my entitlement to your practical conclusion. Having an entitlement to your practical conclusion means that I can trust that you are conducting yourself in a way that does not invalidate your decisions, and if you contravene the practical norms associated with that intention then you let me down in this way.[66]

One might be wondering how this practical authority and entitlement come about. Roth thinks that this authority and entitlement always occurs against certain backgrounds of sociality, but he is rather unspecific about the details. Some authority in these sorts of cases may be established by explicit or implicit consent: you and I could both grant one another permission to make decisions for the both of us. Although consent might be able to establish authority, it's not clear how consent could establish an entitlement to the decision of another as a practical conclusion for me.[67] While Roth doesn't claim to have a complete account of how such authority might come about, he seems to think that such authority is part of our nature

[66] Ibid. pp.390-395.
[67] Ibid.

as social beings, just as it is part of our nature as temporally-extended individual agents that the past deliberations of our past selves enjoy practical authority over our current selves. I take this to be a sort of transcendental argument: as humans, we have the ability to legitimately issue and take up intentions in virtue of our sociality. So, as humans we enjoy the requisite authority and entitlement in virtue of the same.[68]

Roth usually discusses the phenomenon of practical intimacy against a background of collective action that is already occurring. When Kim issues an intention to Jim that he drive first, this is within the context of a larger endeavor (taking a road trip) in which they are already engaging. However, Roth's idea can be quite felicitously extended to the beginning stages of a collective action. If practical intimacy is possible within the context of a collective action, then there is no reason to think that one can't successfully initiate a collective action by issuing an intention whose content includes everyone's roles in some shared endeavor with a simple statement like "Let's take a walk." In fact, this often seems to be how it happens.[69] When multiple agents act on a shared intention that issues from one deliberative source, they perform a collective action.[70] Of course, for this to happen the appropriate practical authority and entitlements will have to be in place. But in favorable circumstances, I think we can often explain the origins of collective action in terms of practical intimacy.

---

[68] This issue is a promising topic for future work.

[69] To adapt an autobiographical example, after a discussion about where we will vacation this summer, one person finally says "Let's just go to Gatlinburg again," and then (despite uncensored grumbling) everyone so intends without further ado. A discussion of this kind is probably itself a kind of collective action, but I don't think that detracts from the point (another example below).

[70] This deliberative source need not be an individual. Plausibly, certain groups (e.g. a committee) can deliberate and issue decisions. Practical intimacy can explain the relationship between such a group and its executors in terms of the roles of issuer and addressee, where the issuer is the group itself and the addressees those individuals selected to carry out the group's will. *see* Roth (2014b) pp.153-156.

A concrete example may help bring out how all this works. Suppose I have previously expressed my desire that you look over my paper, and I've told you that I am free most Wednesdays. One Wednesday you knock on my office door and ask me if I'm busy, to which I reply that I'm not. You say, "Delightful! Let's meet in the common room in a few minutes and talk about your paper." I nod and gather my things. In this scenario, you have intended *for the both of us* that we will meet to discuss my paper now, and hence you have settled the question regarding what I will be doing for the next hour or so via your practical authority. This authority intuitively arises out of factors relating to our circumstance, our relationship, and our nature as social agents. Your issuance and my subsequent adoption make it the case that we share an intention, and this in turn makes our collective action possible. If instead of proceeding to the common room I locked my door and kept working, then I would let you down. You legitimately settled something for me, but without consulting you I flaunted your practical authority by contravening the practical norms associated with your intention. Similarly, if you go to the ice-rink instead of the common room, then you undermine my entitlement to your decision as a practical conclusion for me because you make our meeting in the common room impossible. In doing so, you let me down.

So we have an idea of how the practical intimacy view accounts for collective action. Sometimes, one person has the authority to settle for herself and another what they will together do. If the two act directly on that intention, then they all literally share an intention to carry out the collective action, and this explains the sense in which they are acting *together*. Moreover, when two are settled on a course of action in this way, they are contralaterally committed to one another to perform that action.

Unlike Gilbert's joint commitment, the commitments involved in practical intimacy are integrated with the agent's other practical commitments and reasons. The crux of the practical intimacy view is that your intentions feature into my practical reasoning *just like my own do*. If my reasons against acting on your intention overwhelm the reasons I have in favor of acting (which, according to Roth, will include the reasons *you* had for forming the intention in the first place[71]), then I am not committed to acting. Your intentions bind me, but only defeasibly – just as my own intentions do.

II.V.II. Problems with the Practical Intimacy View

There remain serious questions about the practical intimacy view. The first worry concerns how the practical authority and entitlement come about, and whether the resulting authority structure is compatible with collective action.

In certain contexts, the idea that one person can act on another's intention directly (i.e. without redeliberating about that intention) is relatively uncontroversial. Velleman admits as much:

> There is nothing problematic about first-person-plural intentions in themselves. One person can decide or plan the behavior of a group, for example, if he holds authority or control over the behavior of people other than himself.[72]

Matthew Hanser argues that commands and requests express the issuer's intention that the addressee act in a certain way.

> If the command or request expressed an intention that [the addressee] Φ, then [the addressee] takes on or adopts for himself an intention to Φ. But he does not *decide* to

[71] Roth (2014a) pp.643-647.
[72] Velleman (1997b) p.34.

> Φ. The intention expressed by the command or request was the product of someone else's deliberation and decision, not his own, and he takes the intention directly from this other person.[73]

This is not an unintuitive position; in situations where authority is well established and one-directional, practical intimacy is pre-theoretically plausible. Within a certain circumscribed domain, you really can settle what I will do as my boss or professor. But you might think that this sort of authority isn't really compatible with *shared* intending or a collective action. Velleman says as much:

> …shared intention is not supposed to be a matter of one person's deciding or planning the activities of a group; it's supposed to be a matter of shared intending, in which each member of the group participates equally in forming and maintaining the intention, fully recognizing the others as equal participants.[74]

The suggestion seems to be that one person can intend for another in the way that practical intimacy requires *only* in contexts where the power structure between issuer and addressee is incompatible with the togetherness that we are after in an analysis of shared intention, and hence collective action.

There are a few things that can be said in response to this worry. First, we shouldn't assume that autocratic power structures preclude collective action. Power structures are rarely entirely egalitarian; as such, it would be quixotic to assume that collective action essentially involves complete equality with respect to practical authority.[75] Second, practical intimacy is

[73] Hanser (2015) p.113; at this point, Hanser cites Roth (2004).
[74] Velleman (1997b) p.34.
[75] c.f. Tuomela (2013) pp.160-173.

not incompatible with egalitarianism. It is possible that both you and I have the authority to settle some question, and yet only one of us actually exercise that authority.[76]

There is another problem here that Roth does not address. The practical intimacy view is designed to account for the contralateral commitments between participants in a collective action (at least in small scale cases). But while Roth provides the tools for a plausible analysis of the contralateral commitments between issuer and addressee, he does not say enough about how the contralateral commitments between *all* participants come about when the group is made up of more than two people. If Kim, Jim, and Sarah are going for a walk together, and Kim legitimately issues an intention to Jim and Sarah that they all go walking together (which they then take up), we don't yet have an account of how Jim and Sarah are contralaterally committed to one another. All we know so far is that because they are both addressees, they each have a commitment to Kim and Kim to each of them.[77]

Sarah and Jim are committed to Kim in virtue of her practical authority. Insofar as the content of Kim's intention (which is issued to both Jim and Sarah) is an activity involving all three of them, Sarah is committed to Kim *with respect to* Jim and his activity. This isn't quite the same as Sarah being committed *to* Jim, but I think there are two senses in which she is. First, consider that Kim's practical authority, with respect to this very activity, extends to both Sarah and Jim. As autonomous agents, Sarah and Jim are not just passive objects of Kim's authority; rather, in taking on her intention as their own they become *avatars* of her practical

[76] Roth (2004) p.393-395.
[77] It should be noted that practical intimacy is itself compatible with other analyses of collective action. Nothing about practical intimacy precludes the agreements of the plural subject view, or the sort of intending envisioned by the reductive view. So if we can't draw out contralateral commitments between all parties with practical intimacy, this won't constitute a knockdown objection.

authority. I think that just as Sarah is committed to Kim in virtue of her practical authority, so Sarah is committed to Jim in virtue of Kim's practical authority. Insofar as both Kim and Jim manifest, qua their respective agencies, the practical authority in virtue of which Sarah is committed, Sarah is committed to them both.

Second, although it is true that in this very instance Kim was the one who exercised her authority in issuing an intention, in many instances of collective action *all* parties enjoy a degree of practical authority, accompanied by corresponding practical entitlements, sufficient for them to make decisions for the group. Concomitant with this authority is an ability to settle what it is others will do with respect to an intention that the group is currently executing. What this suggests is that *ceteris paribus* in an egalitarian group, all party members have the authority to legitimately remediate one another's performances by demanding compliance. Kim, Jim, and Sarah are all performing a collective action together, and though it was Kim who issued the intention, all parties have a stake in the game and can legitimately right the ship should things go wrong. For this reason as well, all parties are committed to one another to carry out an intention issued from a legitimate practical authority.

II.VI. Conclusion

The practical intimacy view has some of the best features of both the reductive view and the plural subject view. On the one hand, like the reductive view the plural subject view does not drive a deep wedge between the conceptual framework of individual intentional agency and collective agency. Both imply that one can intend things besides one's own actions (as long as one can settle what's at issue), but neither appeal to mysterious intentions of any sort. Yet practical intimacy is not plagued by problems about settling. On the other hand, like the plural subject view, practical intimacy holds that individuals engaged in a collective action can literally share a single intention. While a shared intention in this sense will be associated with distinct mental states in each individual's head, we can legitimately describe each individual's intention as the same in virtue of the causal and normative origin of those states. Yet unlike the plural subject view, practical intimacy does not have problems with logical structure, and those intentions associated with collective action bind only defeasibly.

While the reductive and plural subject views present interesting accounts of how collective action can come about, they each face serious problems. Practical intimacy gives us a clear and simple picture of how multiple individuals can come to share and act on a single intention. Sometimes, we really do have the authority to settle what others will do, and sometimes others have the authority to settle things for us. When you and I are engaged in a collective action, this sort of shared settling will be a common occurrence. In virtue of this rich exchange across our shared practical perspective you and I are practically committed to one another, and we act together with intentions that we all share.[78]

[78] Thanks to Kevin Falvey, Matthew Hanser, Sonny Elizondo, Sam Zahn, Seyed Mohammad Yarandi, and Catelynn Kenner for helpful comments and discussion. Writing this chapter was certainly a collective endeavor.

III. Directives, Practical Thought, and the Boundaries of Agency, Part 1

III.I. Abstract

In this chapter, I examine directive interactions, which are interactions wherein one person directs, commands, or bids someone to do something. I show that directive interactions are ubiquitous features of social life. I then argue that directive interactions should not be understood as tools for giving someone a special reason to do something. Rather, in a paradigmatic directive interaction, the director issues an intention to the addressee, which the addressee can act upon without independent deliberation. I argue that this way of understanding directive interactions comports well with the function of directives and with the internal experience of directing and being directed.

III.II. Introduction

In this chapter and the next chapter, I am interested in what normally happens when one agent tells another agent to do something and the other agent does it. If you pay close attention to what goes on when humans interact with one another, you will notice that we constantly direct one another in all sorts of ways, many of them subtle, some of them not. I tell you to pass the salt. You tell me to speak up. King Henry orders his soldiers once more unto the breach. And so on. With few exceptions, contemporary philosophers have assumed that telling someone to do something is a way of activating or giving them a reason to act. As a result, discussions of these interactions have mainly been focused on what sort of reason is activated or given by a directive. I think this whole way of understanding directives is misguided. In my view, the main function of a directive is to direct the addressee by communicating an intention to them, which they can act upon directly. As will become clear, the considerations that lead me to reject the predominant way of thinking about directives are mainly rooted in the way in which directives characteristically shape the practical thought of agents issue and accept them.

Interactions wherein one person directs, tells, commands, or bids another person to do something are mostly overlooked in social action-theoretic literature and are relegated to discussions of political, institutional, or organizational authority. These interactions are not taken to be central to our social lives. I think this is a mistake for two reasons. The first is that these interactions pervade many if not most sustained social encounters and are integral to human social functioning. The second is that these interactions reveal something important about the social dimensions of human agency. I will suggest that when studying these interactions it sometimes becomes strained to think of the agents involved as functionally discrete and impermeable spheres of intentional activity. Normally, directing another is a way of extending one's agency interpersonally.

To illustrate what I have in mind, let us begin with an example. Not long ago, my wife called me at our apartment and asked me to find a document for her. My wife had a better idea of where the document was than I did, so she actively directed my search. First she told me to check in the filing cabinet, which I did to no avail. Then she told me to check under the bed, and again I turned up nothing. Finally, she told me to look in a drawer in the coffee table, wherein I found the document. She asked me to scan the document and send it to her, which I did after hanging up the phone.

There are two things we might focus on in this example. If we focus on the whole process of finding the document, we can see that my wife and I were doing something together in a robustly social sense. This brings out the collaborative, joint nature of the process. But if we focus just on my individual actions, such as my checking of the filing cabinet, my rustling under the bed, and my scanning the document, we can see another feature. Insofar as my actions were a product of my wife's deliberative faculties, insofar as I was acting at my wife's direction and guided by my wife's plans and goal, there is an important sense in which my actions were an expression of not just my agency, but of my wife's as well. It is as if my wife's sphere of intentional activity was interpersonally extended such that it overlapped with or circumvolved my own intentional activity.

Or so I will be arguing in this and the next chapter. More concretely, I will be claiming that I was acting on my wife's intention and for my wife's reasons. My actions were, under some description and in a certain way, attributable to my wife as well as to me. This generalizes. One's actions are sometimes done for others' reasons, one's activities are sometimes attributable to others, and one's plans and goals sometimes guide the actions of others whom one controls or with whom one collaborates. One of the upshots of this is that if we want to know what a socially embedded agent is doing, what she is up to, we cannot in every case limit ourselves to what the agent herself is thinking and the ways in which the agent is moving her body and interacting with her environment. We must, on occasion, look to the actions and thoughts of others, not just to provide context for her behavior, but to understand at the most basic level how her agency finds expression in the world.

Here is the plan for this and the next chapter. I begin by presenting several quotidian wherein one person directing another. I then argue that in general we should not analyze interactions of this kind in terms of the giving of reasons, since this analysis is in tension both with the ways that directives seem to influence practical thought and with the relationship between the issuer and the directed action. I endorse a thesis due to the philosopher Abraham

Roth (2003, 2004) called *Practical Intimacy*, according to which agents can take up and act upon another's intention without deliberation, and then I propose that directive interactions characteristically involve the interpersonal transmission of an intention, whereby one agent extends her sphere of intentional activity to include the actions of another.

In the next chapter, I defend the intention transmission paradigm against the objection that it is not acceptable from a psychological or naturalistical standpoint, elaborating in the process on the sense in which a compliant addressee's action is attributable to the agent who issued the directive. I argue that complying with a directive does not inherently involve an abnegation of one's autonomy. Some actions, I suggest, are both autonomous and heteronomous at the same time. Finally, in the last section I conclude the chapter and gesture at some of the broader implications of the ideas defended here.

III.III. The Nature of Directives

We are social creatures. We talk, reason together, develop relationships, coordinate our actions, jointly pursue goals, form collectives, establish institutions, and engage in fellowship for its own sake. One basic feature of our sociality consists in the capacity to get others to do things by directing them. While the exercise of this capacity is more salient in some social contexts than others, it is quite widespread. This capacity makes it the case that agents sometimes have reason to think about what others should do. Consider five examples of interactions wherein this capacity is exercised:

1. Two hikers hiking a trail together come to a fork. One of the hikers is knowledgeable about the trail and determines that they should go left. The knowledgeable hiker turns left, and the other hiker follows.

2. The owner of a lawn mowing service has contracted to mow two lawns over the next two days. One lawn is nearby and the other is out of town. The owner decides that her employees shall mow the nearby lawn today and the out of town lawn tomorrow. She calls her employees, tells them what to do, and the lawns are mowed in the prescribed order.
3. Two housemates are throwing a cocktail party. The housemates deliberate about how best to minimize the impact of the party on their home's cleanliness and mutually decide that they shall have the guests take off their shoes when they enter the house. The guests are told to take their shoes off at the door, which they all do.
4. A married couple's anniversary is approaching. The spouses take turns planning their anniversary celebration. This year, it is the wife's turn. The wife thinks about what she and her husband would enjoy and ultimately decides that they shall have a picnic on the beach and then go dancing. The couple enjoys an evening of picnicking and dancing.
5. A father is taking care of his toddler for the day and decides that they shall go to the zoo together in the afternoon. They go to the zoo and have a lot of fun.

These examples occur in a diverse range of circumstances. What they all have in common is that in each of these examples, an individual or group of individuals thinks about what someone else should do, makes a determination, and then tells (or otherwise indicates to) them to do it. Call interactions of this kind *directive interactions*.

From an external, social psychological perspective there is nothing mysterious about directive interactions. They involve a prosaic type of interpersonal influence or control, which arises whenever humans gather in groups, form hierarchies, or become familiar with one another. If we were only concerned with an external description of directive interactions, perhaps we could leave it at that. But directive interactions involve agents, and agents have points of view that are not wholly describable in external terms. From an internal, agential perspective directive interactions and the influence they involve take a certain form. I think it is worth thinking about the form that directive interactions take for the agents who participate in them, because I believe that this form reveals something interesting about human agency and action.

How, then, should directive interactions be understood from an internal, agential, action-theoretic perspective? One approach is to classify directive interactions as processes wherein the intention of one agent is carried out by another. This is basically the approach I endorse. But despite being a rather natural way of thinking about directive interactions, this is not how directive interactions are typically understood in philosophical discussions of this topic (most of which, incidentally, occur within the literature on political authority). The most common way of understanding directive interactions is in terms of the giving of reasons.

In the next subsection, I discuss this approach to directive interactions. I will suggest that the ways that directives characteristically shape the practical thought of those who engage in them speaks against this approach.

III.III.I. The Reason-Giving Paradigm

According to *reason-giving accounts of directives*, the main function of an authoritative directive is to give an addressee a reason for action of a certain kind and thereby (hopefully) affect the balance of reasons for the addressee.[79] Proponents all agree that directives constitute

[79] Philosophers writing in the literature I am discussing are usually primarily interested in directive interactions that occur in contexts shaped by well-defined, hierarchical authority

or trigger reasons for the addressee to act as prescribed, but they disagree about what these reasons are like.

The most popular reason-giving account is due to Joseph Raz. Raz distinguishes between first-order and exclusionary reasons. A first-order reason is a reason for an agent to perform an action. An exclusionary reason is a reason for an agent to refrain from acting for a reason (Raz 1999 pp. 39-45).[80] According to Raz, directive-reasons are compounds, called 'protected reasons,' composed of both first-order reasons for the addressee to perform an action and exclusionary reasons not to act on certain reasons that bear on that action.[81] Raz's view is motivated by his beliefs about authority's normal function. Authority functions to help addressees act in ways that conform with reasons that apply to them independently of the directive, reasons upon which directives should, in the main, be based (Raz 1986 p. 47).[82] Authoritative directives replace, in the addressee's deliberation, the reasons on which they are based by giving addressees both a first-order reason to perform a certain action (typically, the action that was most supported by the reasons upon which the directive is based) and an exclusionary reason to avoid acting on reasons that the authority has already taken into

relations (e.g. civic, legal, and managerial contexts). I am assuming that for the purposes of action theory directive interactions take the same basic form for participants regardless of the context. This is not to deny that when, say, a directive results in a moral obligation it will be experienced differently by the agent than it would have been if there were no concomitant obligation. Rather, my assumption is that the core of the experience will be the same in both cases.

[80] *Exclusionary reason* is a determinate of the determinable *second-order reason*, which is a reason to act on or refrain from acting on a reason.

[81] The term 'protected reason' was introduced in Raz (1979) p. 18. Technically, directive-reasons are only *typically* protected reasons since there could in principle be directive-reasons that are only exclusionary reasons, e.g. a command to refrain from acting from a selfish desire.

[82] This is tied to Raz's view about how authority is normally justified. Authority is normally justified insofar as the addressee would in fact better conform to reasons that apply to her independently of the directive by trying to follow the authority's directives than by not doing so, provided it is more important that the addressee conform to reason than decide for herself in the relevant domain (Raz 2009b pp. 136-137).

consideration in issuing the directive (or, at least the contrary ones[83]) (Raz 1986 pp. 58-59). Directive-reasons "preempt" (i.e. replace) an addressee's judgment about some of the directive-independent reasons by supplying a pre-packaged protected reason that reflects those directive-independent reasons, effectively simplifying an addressee's deliberative burden by making it the case that he has fewer reasons to consider. [84]

On reason-giving accounts like Raz's, directive interactions do not involve anything like one agent acting on another agent's intention. Authorities decide whether to issue directives, which give addressees reasons. Addressees decide what to do in light of these reasons. In both cases, it is more or less business as usual: the agents form and act upon practical judgments concerning what to do given the reasons they think apply to them. Despite its widespread acceptance, I do not think the reason-giving paradigm provides action theorists with the best framework for thinking about directive interactions. First, reason-giving accounts, even those as sophisticated as Raz's, fail to adequately capture the distinctive way in which directives normally shape practical thought for addressees. Second, they obscure the issuer's perspective on, role in, and connection to the activity of a compliant addressee by partitioning the contribution of the issuer and the addressee into discrete parts.

Let us consider how directives shape practical thought. It will be helpful to begin by considering some of the reasons why the practice of using directives exists and relatedly why individuals participate in this practice. One reason the practice exists, which Raz is keenly aware of, is that directives enable groups of individuals to efficiently utilize practically-

[83] To my mind, Raz is not always clear about whether directives exclude only contrary reasons or all reasons which the authority considered in her deliberations.
[84] There are many other reason-giving accounts, which vary across several dimensions. For instance, see Friedman (1973), Finnis (1980), Hart (1990), McMahon (1992), Korsgaard (1996), Darwall (2009), and Enoch (2014).

relevant knowledge that is unevenly distributed within the group. This property of the practice of using directives also explains why individuals choose to participate in the practice. For instance, consider the example with which I began this chapter. My wife actively directed my search because she had a better idea about where the document was than I did. Rather than communicating that information to me so that I could conduct the search autonomously, it was more efficient for my wife to simply make a determination about what I should do and communicate that determination to me directly. By acting on this determination without considering the reasons myself, I was able to utilize my wife's practical knowledge without having to go through the trouble of acquiring and assessing it myself.[85]

A closely related point is that directives can consolidate and reduce the deliberative burden facing a group. Deliberation takes time and energy. When agents are pursuing a goal, they typically have to deliberate about the means required to achieve that goal. But when the issuer of a directive is in a position to calculate the required means and communicate that calculation in the form of a directive, directives can reduce or eliminate the need for the addressee to engage in means-end deliberation. This possibility also helps explains the existence of the practice of using directives and why agents are motivated to participate in that practice.

Agents are motivated to employ and accept directives for other reasons. Consider case (2), which involved the lawn mowing operation. Why is the owner of the company in a position to tell her employees which lawns to mow? The short answer, of course, is that the owner and employees have consensually entered into a contractual relationship, which makes it appropriate for the owner to direct the employees. But what makes this particular contractual

[85] The motivations of the agents in (1) and to some extent case (5) to utilize directives can also be explained in this way.

relationship apt in this case? Why is the relationship such that it is the owner who is giving orders rather than the employees?[86] The answer, I think, has to do with the fact that the owner and employees are engaging in a joint enterprise in which the owner is specially invested. The owner's investment has monetary and legal components: the owner's money is directly tied up with the success of the company; she owns the company, equipment, and so on. But there is more to it. The company is the owner's creation, a product of her designs. It is, basically, a mechanism through which the owner expands the reach of her agency in the world. She can mow more lawns and get more money with a team than she can by herself. The company is not a mechanism of this sort for the employees, and this marks an asymmetry in the employer's and employees' relations to the joint enterprise in which they are all engaged. The owner is the source of the company's purpose, its prime mover, as it were. This is the other component of the owner's special investment, which explains why it is *her* enterprise. And it is one of the conditions that make the authority relations apt.

The preceding discussion is idealized, of course. Not every business owner needs to think of things in this way, and anyway the class of businesses and related enterprises is very heterogeneous. But this discussion captures something of what it means to say that one agent is working *for* another. One agent borrows or co-opts another's agency for her own purposes. In these conditions, agents often feel that directives are appropriate. And the practice of using directives likely exists in part to enable these sorts of relationships.

Perhaps one of the most important explanations of the practice and one of the most common motivations that agents have to employ and accept directives on a day-to-day basis is

[86] This question might seem to rest on a conceptual confusion, but it does not. There are many cases wherein employees direct their employers. For instance, personal trainers—or at least certain types of personal trainers—are expected to regularly direct their employers (i.e. clients) to do things (work out more, eat healthier, etc.).

that directives facilitate smooth interactions between agents who are cooperating, assisting one another, or acting jointly. A mutually recognized entitlement to issue directives can drastically reduce the communicational friction between cooperating (or potentially cooperating) agents. The fastest way to get you to pass the salt at the dinner table is to simply tell you to pass the salt. The fastest way to resolve the coordination problem about who will cross the one-way bridge first is to tell you who will cross it first. And the fastest way to get the project of cleaning the house started is to assign you tasks. The ubiquity of directives is to some extent masked by the fact that directives are not always (and in some contexts are usually not) communicated using the imperative mood. Sometimes, a question (e.g. "Can you pass the salt?"), an assertion (e.g. "I need the salt."), or a gesture (e.g. pointing to the salt) will do. Despite the superficial differences between these communicative acts, typically they all function as directives when they occur in a context wherein there is a mutually recognized presumption that addressees will respond by performing the indicated action.[87] It is tempting when one focuses on extreme cases ("once more unto the breach!") or legal cases ("stop at the stop sign!") to think that directives are normally experienced by adult agents as interloping demands that mandate willful renunciation of self-governance and self-interest. But the vast majority of directives are not experienced in this way and mandate nothing of the sort. Especially when one is immersed in a cooperative task with familiars, directive interactions are typically so fluid that they are often not even registered as directives in conscious thought. As Christine Korsgaard puts it, it

[87] The extent to which such presumptions characterize any given relationship or interaction is an empirical question. There are obvious cases. If you walk up to a stranger on the street and ask for five-hundred dollars, there is obviously no mutually-recognized presumption that the stranger will comply. But if you ask a stranger on the street to give you directions to a well-known local landmark, there probably is. This is a contingent sociocultural fact, of course. Notably, a presumption of this kind can be counteracted. For instance, you could say, "Would you mind directing me to the beach? If you're too busy, don't worry, it's not a big deal."

is very hard *not* to "respond with the alacrity of obedient soldiers" to the directives of others (1996 p. 141). And this is part of what enables directives to play the social lubricative role that they do.

To sum up, directives are employed and accepted for various reasons. They are tools for effectively utilizing practically-relevant knowledge that is unevenly distributed across a group of people. They can be used to consolidate deliberate costs. They are appropriate when an agent involves or co-opts others in her projects. And they facilitate smooth interactions between agents who are doing things together.

The types of situations in which directives are typically employed give us some indication as to how directives characteristically shape the practical thought of addressees. When an addressee is directed to do something routine by someone he recognizes as able to do so in a context where such directions are common, accepted, and expected, the addressee will characteristically adopt the directive automatically as a constraint on his practical reasoning and future actions. This is not a form of temporization. The directed action, much like one's intended actions, is presented to the addressee as something to be done, not as something to be considered or deliberated about. The ability to take directives as exogenous constraints in practical thought is advantageous to social creatures like us, since otherwise directives could not function effectively in the ways that they do.

To see why not, consider what it would be like if we could only give one another advice but could not direct. It can sometimes be difficult to distinguish a giving of advice from an issuing of a directive, but when it is clear that it is advice rather than a command that is being given, advice does not present action to the addressee as to be done. Advice presents action as *to be considered*. It is still up to the advisee to deliberate about what to do given this advice, and this deliberation might be quite difficult or time-consuming. Sometimes advisees can, of course, (rationally) respond to advice by more or less automatically adopting the advised course of action, but this is only possible when the advice makes it abundantly clear to the addressee that the suggested course of action is the correct one. Many situations wherein advice or directives are apt are not like this. Because advice leads to action only through the advisee's deliberative channels, and hence through the advisee's practical perspective, advice cannot be used to facilitate social interactions or consolidate deliberative costs as effectively as directives can. If my wife advises me to look in the drawer, it becomes necessary for me to determine how best to incorporate that advice into my own thought about what to do given all the information available to me. Directives normally eliminate this requirement by circumventing my deliberative channels.

So, directives work their magic on addressees insofar as they present actions to addressees as *to be done* rather than *to be considered*. This suggests a problem for reason-giving accounts. Reasons do not control action directly; the connection between reasons and action is mediated by rational, reason-responsive processes of agents. When reasons push in different directions, they must somehow be agglomerated by an agent who has the power and wherewithal to adjudicate between their competing claims. And even when an agent recognizes that all the reasons point in a single direction, the agent must take the psychological step of deciding to act in light of those reasons.[88] Herein lies the problem. If directives merely function

[88] This decision need not be preceded by ratiocination. The point is that there is a real psychological difference between recognizing that one has overwhelming reason to do something and being settled on doing it. Michael Bratman makes basically this same point while arguing against analyzing intentions in terms of belief-desire pairs. He suggests that one can both have a predominant desire to ϕ (and, hence, have overwhelming reason to ϕ) and believe that as a result one will ϕ without being settled on ϕing and without viewing the issue of whether to ϕ as given in deliberation. For that, a decision (and a resulting intention) are required (1987 pp. 18-20).

to give addressees reasons, then they require the agent to take a psychological step that involves drawing upon the full resources of the agent's own practical perspective: his relevant background beliefs, his beliefs about the reasons that apply to him (including the directive-reason), his own methods of reasoning, and so on. And this is the very psychological step that directives often function to circumvent.

One might object to this by pointing out that agents normally accept directives because they take themselves to have reason to do so. Moreover, agents can (and often do) deliberate about whether they should be disposed to accept directives from a particular source. Agents even sometimes deliberate about whether to act in accordance with a particular directive. So, there is an undeniable sense in which directives work their magic through the addressee's deliberative channels, and this calls into question whether directives really present actions as *to be done* rather than *to be considered* and whether there really is a problem with the reason-giving approach.

In responding to this objection, we should first note that when an agent deliberates in advance of some directive about whether to accept directives from that directive's source, the directive itself need not (and typically will not) feature into that deliberation. If an addressee decides to accept directives from a particular source and as a result becomes disposed to do so, then a directive issued from that source will normally cause the addressee to act without any further deliberation. The fact that the addressee's deliberation is a precondition of the addressee's acceptance of the directive does not show that the directive affects the agent's actions by going through the agent's deliberative channels, since the addressee's deliberation does not intervene between the directive and the action.

Moreover, even if the deliberation of an addressee does intervene between a directive and the addressee's action, there is sometimes still an important sense in which the directive functions to circumvent the addressee's deliberative channels. To see why, we should distinguish two different sorts of case. In one sort of case, an addressee is directed to perform some act-type φ, and the addressee deliberates about whether to φ. In this sort of case, the directive does not circumvent the addressee's deliberations but rather moves the addressee to act via the addressee's deliberation, much like advice. But in another sort of case, the addressee is directed to perform some act-type φ, and the addressee deliberates about whether to comply with the directive, where the content of the directive does not feature into to the addressee's deliberation. The latter type of case is the cotemporaneous analogue of the situation where the addressee deliberates in advance about whether to accept directives from a certain source. And in this type of case, the directive causes the addressee to φ without the addressee's deliberating about whether to φ (under the φ description). So, the directive still functions to circumvent the addressee's deliberation about φing, even though it does not circumvent deliberation altogether.

What all this shows is that deliberation about directives need not preclude the use of directives to circumvent deliberation. And just because I have indirectly deliberated about whether to comply with a directive does not mean that the directive presents itself to me as *to be considered* rather than *to be done*. Thus, the reason-giving account faces a problem.

Of all the reason-giving accounts on the market Raz's view probably comes the closest to capturing the distinctive way that directives influence practical thought, since according to Raz directive-reasons function to preempt the addressee's practical judgment about the reasons on which the directive is based. But the reasons on which any given directive is based need not (and, I think, normally will not) include all the reasons that apply to the agent at the time of action, meaning that the exclusionary reason associated with a directive need not exclude all the non-directive reasons that apply to the agent at the time of action. In cases like this, the task of determining what to do given all the reasons that apply to the agent at the time of action will be non-trivial. And even when a directive-reason clearly does make it all-things-considered rational to act as prescribed, the agent must still draw upon deliberative or reason-responsive processes in order to make a determination that this is the case. Perhaps Raz's account fits the phenomenon of advice better than the phenomenon of directives. Arguably, advice functions to give an advisee a better idea about what some of the reasons that apply to him call for; directives function (at least some of the time) to remove the need for independent deliberation altogether.

So, one reason why the reason-giving paradigm does not provide action theorists with the best framework for thinking about directive interactions is that it does not adequately reflect the way in which directives shape the practical thought of addressees. The second reason is that it misrepresents or obscures the issuer's perspective on, role in, and connection to the activity of the addressee. Consider again the case of the lawn mowing operation. I suggested that part of the reason it seems appropriate for the participants in that enterprise to enter into a relationship such that the owner is the one who is in a position to give orders has to do with the fact that the owner is co-opting the employees' agency for her own designs and not the other way around. This case is idealized. Still, it is clear that we often do co-opt others in our projects in this manner. And if co-option like this is possible, it must be possible for an agent to do more than just advise others. If the project is to remain mine rather than simply ours, it must be possible for me to do more than just tell you what I think the reasons that apply to you recommend that you do.

An agent who is in a position to direct an addressee normally sees the actions and agency of the addressee as being at her disposal in deliberation. The addressee is incorporated into the prospective possibilities that are the potential objects of choice for the deliberating agent, and the addressee's capabilities partly define the courses of action available in deliberation to the agent. An agent who can direct an addressee to do what she wants will not normally deliberate about whether to *get* the addressee to act as she desires, nor will she deliberate about *how* to get the addressee to so act. She will simply deliberate about whether the addressee should (or shall) act in that way.[89] The addressee is, in a sense, incorporated into the agent's first-person practical perspective. For example, the company owner may simply deliberate about which lawn her employees will mow tomorrow, since her employees are at her disposal. Or consider my wife's perspective on our search for the missing document. My wife's practical reasoning did not focus on what to say to me or on how to get me to act so that the document would be found. Her reasoning was focused on where I should look in order to find the document. My activities became the object of her deliberation, and this was possible because of her entitlement to direct me and her expectation that I would comply.

It is more difficult to make sense of the inclusion of the addressee into the issuer's first-person deliberative perspective if issuing a directive simply amounts to giving an addressee a reason for action. If your directive merely gives me a reason to act in a certain way which I

[89] This does not imply that agents do not need to do something in order to direct addressees. It only implies that in many situations one does not need to deliberate about the relevant means. Similarly, I might deliberate about whether to tell you that some proposition is true without deliberating about *how* to tell you, but this does not imply that I do not need to employ means to tell you (or that telling you is a basic action for me (under that description)).

must, as a rule, weigh according to my own scales, then it seems you must treat me as if I fall squarely outside the sphere of your own agency and intentional activity. Insofar as you conceptualize a fundamental distinction between your own agency and the world, I am placed within the world, a thing to be moved as a means to your ends, not a thing whose activities are an object of choice for you. By analogy, compare your perspective on me if in order to get me to act you must first attach an incentive to the desired action or persuade me with arguments. In situations like these, my actions are not potential objects of choice for you. Rather, my actions are possible outcomes of your actions (and thus of your potential objects of choice).

One might object by claiming that the inclusion of an addressee into an agent's first-person deliberative perspective only requires that the agent be confident that the addressee will do as he is bid. Consider for the purposes of analogy a skilled operator of a telemanipulator. A telemanipulator is a mechanical device with which an operator can manipulate hazardous objects at a distance. A skilled operator of a telemanipulator can perfectly well deliberate about whether to pick up this or that with the telemanipualtor rather than about whether to make the telemanipulator pick this or that up, and this seems not much different than the way the owner of the lawn mowing company thinks of her employees in the process of deliberation. The operator can think of the telemanipulator in this way because he realizes that the telemanipulator is likely to respond appropriately to his skillful inputs, and by analogy, the objection goes, the reason why directors normally see the actions and agency of addressees as being at their disposal in deliberation is that directors can be confident that addressees will comply with their directives. So, the claim that reason-giving accounts are in tension with the way that agents think about addressees is false, since an agent can give an addressee a reason while being fully confident that the addressee will act on that reason.

The analogy is undermined by the fact that addressees but not telemanipulators have their own independent practical perspectives. Telemanipulators do not have beliefs, desires, or intentions. They do not assess reasons or shape the world in their image. They are tools that mindlessly bend to the skillful manipulator's will and practical judgment. But when one is dealing with an agent that one must manage by giving reasons, one is forced to think of the agent's practical perspective as fundamentally distinct from one's own. I must give the agent reasons because what the other agent does depends not on how I see things but on how she sees things, and yet she sees things differently than I do. Fortunately, reasons have a public[90] (or transpersonal[91]) character; what counts as a reason for one agent will generally count as a reason for another, and so the gap between practical perspectives is often bridgeable. But still there is a gap, which is recognized and appreciated by both participants in a reason-giving interaction.

The argument on the last few pages is designed to show that the reason-giving paradigm is in tension with the way issuers think about the agents they direct. The connection between an issuer and an addressee is tighter than the connection between a reason giver and a reason receiver, and this is reflected in the way that issuers deliberate. The final consideration I want to proffer in order to unsettle the reason-giving paradigm before I present my own view relates to the connection between an issuer and a compliant addressee's action.

Normally, if you persuade, advise, or otherwise do something that gives me a reason to act in a certain way and I do act in that way, you are not responsible for my action, and my action is not attributable to you in anything like the way that your own actions are.[92] This is

---

[90] See Korsgaard (1996) pp. 134-142.
[91] Burge (2013) p. 392
[92] You may, of course, be causally responsible for my action. But this sense of attributability is much weaker than the sense I have in mind here (your biological parents are causally

not to say that you are not responsible for influencing me as you did. But we typically think that so long as I make my own decision about what to do and my freedom is not constrained in any way, I alone am responsible for my actions.

Things are quite different if you command me to do something and I comply. We typically think that issuers of directives are responsible for the actions of compliant addressees. Suppose a CEO knows that her company's profits are dropping but in order to deceive potential investors decides to have one of the company's PR representatives announce that the company's profits are increasing. The CEO is morally responsible for the PR representative's announcement. And there is a sense in which the PR representative's announcement is attributable to two people--both the PR representative and the CEO. This is an interesting fact about directive interactions that is not acknowledged frequently enough. An action's being attributable to multiple individuals is not particularly unusual, since joint actions are usually attributable to multiple participants. But the PR representative's announcement is not obviously a joint action.[93] Arguably, the PR representative's announcement is an *individual*

---

responsible for all your actions in that very attenuated sense). See Subsection 3.2 for a more extensive discussion of the various senses in which an action can be attributable to an agent.

[93] One might have the intuition that the announcement clearly is a joint action. Certainly many directive interactions occur within the context of joint action (like when my wife and I were searching for the document). But in this particular example, I am imagining that the CEO did not contribute anything to the announcement except to issue the directive. Without taking a position here on what exactly a joint action is, I propose that this is not enough of a contribution to make the action truly joint (let us assume that the CEO never checks to see if the directive is carried out or even thinks of it again). If this proposal seems unconvincing, then it may help to imagine that the CEO dies of a heart attack immediately after sending the directive, several days (or weeks) before the announcement. In this revised scenario, it seems at the very least odd to say that the CEO and the PR representative are jointly making the announcement when at the time of the announcement the CEO is in an urn. And yet, even in this revised scenario we are inclined to say that the CEO is responsible for the announcement, because in a morally relevant sense it can be attributed to her.

actions that is attributable to multiple individuals.[94] And this is why it is not infelicitous to say that the CEO made an announcement (through her PR representative) for which she is responsible.

The fact that a compliant addressee's action is attributable to and is the responsibility of the issuer (in addition to the addressee) is reflected in the way that issuers typically view these actions. It is normal for an issuer to associate herself with the addressee's action, and this association informs the practices surrounding directive interactions.

To illustrate this, I am going to borrow an idea from Tyler Burge's essay "Reason and the First Person" (1998/2013). Sometimes when one rationally evaluates an attitude or action, it is not *ipso facto* rational to attempt to implement that evaluation. For instance, suppose you see me attempting to persuade university administrators to pay graduate students more by giving a speech in the university quad, and suppose you reason that I should stop wasting my breath since I have selected the wrong location and my speech is unpersuasive. Even if you are right, it is not necessarily rational for you to implement your evaluation by telling me what you think or halting my speech. Since we have different practical perspectives (including different information), I might not see it that way, and you might not be able to convince me. Your evaluation of my action, then, does not provide you with an *immediate* reason to implement that evaluation. You need to engage in further reasoning (i.e. further evaluation) to determine if it is rational to implement that evaluation.

[94] This might sound paradoxical, but it is not, since the sense in which the PR representative's action is attributable to the CEO is not the same sense in which the PR representative's action is attributable to the PR representative. I discuss the relevant sense in Subsection 3.2.

Sometimes, though, the rational evaluation of an attitude or action does give one immediate reason to implement that rational evaluation.[95] Burge argues that first-person concepts (and only first-person concepts) function to mark conceptually those attitudes or actions which one is immediately (but possibly defeasibly) required to modify, maintain, or bring about in light of one's rational evaluations of those attitudes or acts (ibid. pp. 391-392). If *I* come to recognize that the speech is ineffective and should stop at once, then I am *ipso facto* rationally required to stop the speech. The way that I conceptually mark the immediate relevance of my own evaluations to the speech is by associating first-person concepts with the speech--i.e. by thinking of it as *my* speech.

I want to claim that issuers' rational evaluations of compliant addressees' actions sometimes give issuers immediate (possibly defeasible) reason and motivation to implement those evaluations by modifying or supporting the evaluated actions. When this is true, and when there are no defeaters, issuers do not need to reason about whether to implement their evaluation; it is *ipso facto* rational for them to do so. For instance, suppose that you are the president of the graduate student union and you direct me, a union steward, to give a speech in the quad in order to persuade administrators to raise graduate student wages. I start my speech, but you soon realize that the speech is having the opposite of the desired effect; the time is not right, and as a result the speech is likely to galvanize opposition to the wage increase. You determine that, all-things-considered, the speech should stop. Your evaluation of my speech is immediately rationally relevant to the question of whether you should intervene. No further reasoning is needed for it to be rational for you to implement your evaluation by stopping the

[95] If this were not the case, then rational evaluation could never lead to a change in attitude or action, since every evaluation would require yet another evaluation before it could lead to action (ibid. 391).

speech. Not every evaluation you might make of my speech is like this. If you think that I pronounce the word ‘proletariat’ in a weird way, you may not *ipso facto* have reason to implement your evaluation. But at least some of the time, some of issuers’ evaluations of compliant addressees’ actions seem to have immediate rational import.[96]

Burge would not be entirely happy with this. He thought that rational evaluations of others’ attitudes or actions could never give one immediate reason to implement those evaluations, for two reasons. The first is related to the fact that evaluations are dependent upon the intentional systems in which they occur. An evaluative judgment that is reasonable relative to my beliefs may not be reasonable relative to yours, and Burge thinks it follows that no (non-self-evident) evaluation I make of your action gives me immediate reason to implement that evaluation (ibid. 392-393). The second reason has to do with the fact that implementing an evaluation transpersonally always requires the taking of means. To affect your attitude or activity, I must think about how to affect you, and this means that an intervening practical premise must mediate between my evaluation and the implementation of my evaluation. Burge asserts: “One can propose to [modify another’s attitude or act] only by force or persuasion. One’s power over, and responsibility for, the attitude (or activity) are not direct” (ibid 393).

I think that Burge’s worries about the immediate transpersonal applicability of rational evaluations generally only apply when the object of evaluation falls outside of the sphere of intentional activity in which the evaluation occurs. Events that embody an agent’s plans and projects are part of that agent’s sphere of intentional activity. Since directives are a way of including another in one’s plans and projects, directives extend one’s sphere of intentional

[96] Context probably plays a large role in determining what sorts of evaluations give issuers immediate reason to intervene with an addressee’s action. I suspect the judgment that the addressee’s action, despite being performed adequately, is not achieving the issuer’s end usually gives the issuer an immediate reason to implement that evaluation by intervention.

activity to include the actions of others. Since one's sphere of intentional activity does not correspond perfectly to the boundaries of one's body and mind, and since others' actions can occupy that sphere, one's evaluations can have immediate transpersonal applicability. When you direct me and when I comply, your evaluations on certain matters (e.g. should I continue doing this?) matter at least as much and at least as directly as my own evaluations of those matters do. And while implementing your evaluation involves doing something else, Burge is wrong to think that you can only modify my action by "force or persuasion." Simply telling me to stop is enough. As I suggested above, in many cases agents who are in a position to issue directives do not need to think about how to effect that directive at all; rather, they simply take it for granted in practical reasoning will lead to action.

So, issuers' evaluations of addressees' actions sometimes have immediate rational import. And I take it that this is a fact that participants in directive interactions sometimes recognize. If Burge is right about the role of first-person concepts in conceptually marking those attitudes or actions for which one's rational evaluations have immediate relevance, then issuers sometimes associate their first-person concepts with addressees' actions. In other words, issuers sometimes think of the addressee's action as *my* (or *our*) action. This amounts to an acknowledgment on the part of the issuer of "proprietary power over, and responsibility for" the addressee's action (ibid. p. 394) as well as a recognition that the action falls within the issuer's sphere of intentional activity.[97]

[97] It may be worth noting in connection with Burge that Burge's social externalism about mental content (1986) seemingly commits him to the view that a person's intentions can be determined by others in a way that does not involve directives at all. For the content of an intention can be shaped by one's linguistic community. For instance, suppose I intend to make a spaghetti dinner but do not realize that 'spaghetti' denotes a particular type of noodle. The content and success conditions of my intention have something essentially to do with spaghetti noodles, and this is due to the practices of my linguistic community, not to some feature of my individual psychology. Moreover, when I am boiling macaroni noodles with my intention, it

I have argued that normally the actions of compliant addressees are attributable to, and the responsibility of issuers (in addition to addressees) and that sometimes (perhaps even normally) issuers think of addressees' actions as their own insofar as they take their evaluations of those actions to have immediate rational import. To my mind, none of this makes much sense if the main function of a directive is to give addressees reasons. Apart from cases involving unusual coercion, deception, ignorance, gullibility, or obsequiousness, someone who persuades, advises, or otherwise does something that gives another a reason to act in a certain way is not normally responsible for what the other does with that reason. An issuer's normal perspective on and connection to an addressee's action makes sense only if we take seriously the idea that directives are a way of involving others in one's own intentions, plans and projects.

My discussion in this subsection has centered on the distinctive ways that directives shape the practical thought of the agents who issue and receive them. My overall strategy has been to show that reason-giving accounts of directives are in tension with what directive interactions are like for the agents who participate in them. I have suggested that from the perspective of addressees, directives normally resemble intentions more closely than reasons. And I have suggested that from the perspective of issuers, addressees and their actions are to some extent incorporated into the issuer's first-person practical perspective. In the next subsection, I discuss and endorse an alternative view of what is happening in directive interactions that I believe better comports with these data. On this view, complying with directives normally involves carrying out other agents' intention.

seems right to say that I am in fact making a spaghetti dinner but doing it wrong. Thanks to Sonny Elizondo and Chris Britton for discussion about this point.

III.III.II. The Intention Transmission Paradigm

Reason-giving accounts of directives start off on the wrong foot. Directives enable one to include others directly in one's intentions, plans, and projects. When an addressee complies with one's directive, one's intentions, plans, or projects are thereby carried out. The interpolation of reasons into the center of this process obscures these simple truths. For this reason, we should reject reason-giving accounts.

I endorse the view that directive interactions normally involve a form of interpersonal intention transmission, whereby one agent, the issuer, makes a decision and settles a practical question about what another agent, the addressee, shall do. To make sense of how this is possible, I am going to adopt the thesis, originally due to Abraham Roth (2004), that intentions can be transmitted from one person to another. Just as one's own decisions are preserved across time via intentions, decisions are preserved across persons via intentions. So, to communicate a directive is to communicate one's intentions for someone else, and to take up and comply with a directive is to take up and act on another's intention. The rest of this section is dedicated to defending, clarifying, and elaborating on this view as well as exploring some of its wider implications for the nature of agency and social interactions.

The place to begin is with Abrahm Roth's *Practical Intimacy* thesis. Normally, one acquires an intention by deliberating and making a decision. Suppose I am preparing for a hiking trip. I have time to do only one thing I want to do: Either I can break in my new hiking boots or buy a walking stick. It occurs to me that while I can find a walking stick on the trail, my boots will give me blisters if they are not broken in beforehand. I reason that the best way to break in my boots is to go on a long walk while wearing them, and so this is what I decide to do. My decision results in an intention to go on a walk with my boots on, which preserves the psychological and normative force of my decision. My decision is consummated by my intentionally walking in order to break in my hiking boots.

In addition to this standard way, Practical Intimacy says that it is possible to acquire and act upon an intention without deliberating, consciously or otherwise, by taking up an intention that has been transmitted interpersonally (Roth 2004 p. 383). Practical Intimacy presupposes that it is possible both to form an intention whose content includes someone else's action (e.g. I intend that she run) and to communicate such an intention (e.g. by saying "I intend that you run").[98] Call the person who forms such an intention and successfully communicates it to the distinct person who features into the content of that intention *the issuer*, and call that distinct person *the addressee*. According to Practical Intimacy, the addressee can take up the issuer's communicated intention and act on it as if it were her own without "re-issuing" it – i.e. without independent deliberation.[99]

When an intention is transmitted interpersonally in the way that Practical Intimacy describes, the issuer and addressee come to share an intention with the same content and deliberative source.[100] This has important consequences for the teleological structure of the addressee's action, namely the reasons for which she acts, the descriptions under which her action is intentional, and the relevant standards of success. Ordinarily, the teleological structure of my action is determined by my intention and the practical reasoning behind it. I went on a long walk with my hiking boots because I wanted to break in my boots; I was intentionally

---

[98] This presupposition will be defended in below

[99] Practical Intimacy has analogues in the philosophy of testimony. Some philosophers in that literature argue that testimony preserves beliefs across persons just as semantic memory preserves beliefs across time (Burge 1993, Dummett 1994; cf. Roth 2004 n. 54, 56 pp. 407-408).

[100] There is a question as to how literally one should understand this sharing talk. I address this question below.

breaking in my boots; and if in walking I had not broken in my boots, my action would have been a failure. My action would not be for the sake of these reasons nor subject to these standards if my intention were not related to my action in a special way.

Practical Intimacy says that the issuer's intention, and by extension the issuer's practical reasoning, can be related to the addressee's action in just this special way. More precisely, when agent X takes up agent Y's intention and acts on it, the teleological structure of X's action is determined by aspects of Y's practical reasoning in much the same way that the teleological structure of X's action is determined by aspects of X's own practical reasoning when X acts on X's own intentions. In Roth's terminology, X is *entitled* to Y's reasons for intending, even if X is not attending to or aware of Y's reasons (Roth 2014a pp. 643-645; Roth 2017 pp. 82-86).[101] For example, suppose I am birdwatching with a friend. My friend spots a goldfinch in a tree and, so that I can see it, tells me to look that direction. I take up that intention and act upon it. In doing so, I am looking at the tree not (or not just) because my friend told me to. I am looking at the tree for whatever my friend's reasons are, i.e. so I can see the goldfinch. If I fail to see the goldfinch, then my action is a failure. In this way, I am entitled to my partner's reasons, which determine the teleological structure of my action. There are limitations, of course. If I could not understand the reasons, if I would not or could not endorse them, then I am not entitled to my partner's reasons (Roth 2014a p. 646, 2017 pp. 86-90). But sometimes another can decide for me what I will do in such a way that my actions are an extension of their practical reasoning, not mine.

[101] Notice that in the ordinary case, my own reasons can determine the teleological structure of my actions even if I am not attending to, or have temporarily forgotten, why I am doing what I am doing (Roth 2014 p. 644; Hanser 2015 pp. 116-117; Roth 2017 p. 78).

Practical Intimacy provides us with a basic framework for understanding how directives work.[102] Normally, directing another is a way of expressing one's intentions for the other. Normally, complying with a directive involves taking up that intention, in the way Practical Intimacy suggests, and acting upon it. And normally, the teleological structure of a compliant addressee's action is at least partially determined by the content of the issuer's deliberations. While I will discuss some possible exceptions to this rule below, I think this proposal comports with the way directives characteristically shape practical thought and the way we tend to think about the actions of compliant addressees.[103]

Recall that directives are used to effectively utilize unevenly distributed practically-relevant knowledge, to consolidate deliberative costs, to co-opt others in one's projects, and to lubricate the interactions of cooperating agents. Intentions are attitudes that constrain deliberation and guide conduct. Absent a relevant change in internal or external circumstance, an agent who intends to ϕ at t_1 will not normally deliberate about whether to ϕ at t_1. And when t_1 comes, the agent will normally ϕ straightaway. Directives trivially consolidate deliberative costs if they involve interpersonal intention transmission.[104] Since intentions encapsulate the knowledge that was drawn upon in the deliberations that produced them, addressees can utilize the issuer's knowledge directly by taking up and acting upon the issuer's intention. There is no

[102] To my knowledge, Roth does not make the general claim that directive interactions should be understood in terms of Practical Intimacy. His main motivation for adopting Practical Intimacy seems to be the work it can do in helping to explain the source of the intragroup norms that govern the actions of agents engaged in joint action (Roth 2004 pp. 390-395; for a discussion of these norms, see Gilbert 1990).

[103] To be clear, I am not claiming that all instances of interpersonal intention transmission involve directives. The former may be a broader phenomenon. For instance, Roth (2016) argues that when one accepts a promise, one issues an intention to the promisor that the promisor acts as promised.

[104] Directives will not necessarily *eliminate* deliberative costs, however. The issuer's intention may be very general, and even if it is quite specific, possibly an addressees will need to engage in means-end reasoning in order to carry out the issuer's intention.

more straightforward way of co-opting you in my plans and projects than to literally issue my intention to you so that you can act on it yourself. And if you and I can dynamically exchange intentions with ease, then our interactions are likely to become more fluid and our agency more unified as the boundaries that separate our activities become less distinct.

The proposal I prefer also helps us make sense of the director's perspective on, role in, and connection to the addressee and his action. I suggested that the actions of compliant addressees are typically attributable to issuers. The PR representative's announcement is also, in a sense, the CEO's announcement. It is easier to understand the sense in which this is true once we recognize that the PR representative was carrying out the CEO's intention as opposed to simply acting as he thought best given the reasons that he believed applied to him. For the representative's action is the natural terminus of the process whereby the CEO's intention materializes in the world.

I also suggested that normally agents who are in a position to direct addressees incorporate addressees into their first-person deliberative standpoint and think of some of the addressee's actions as their own. The company owner can deliberate directly about which lawn her employees will mow tomorrow; her employees' potential actions become the objects of her deliberations. And when the employees do mow the lawn the owner selects, the owner associates herself with that activity. She is answerable for it; it is performed for her reasons. All of this is explicable if we assume that issuers can deliberate and form intentions for addressees, which can then be transmitted to them and acted upon.

One might hesitate to accept that the teleological structure of a compliant addressee's action is determined (at least in part) by the deliberations of the issuer. Because intentions determine teleological structure by preserving the psychological and normative force of the deliberations from which they issue, this is not a detachable feature of the view. But I think upon reflection it is an intuitive one. We recognize that a commander's reasons partly explain the purpose of, and set standards for, an addressee's compliant action. This is true even when those reasons are not endorsed by the addressee.[105] For example, consider a soldier who is disposed to comply with his commander's directives only because he wants a paycheck. Should his commander direct him to perform a task – say, guard a door – that is crucial for some larger objective the commander is pursuing, then it makes sense to say that his action has the purpose of contributing to the larger objective and that his action is a failure to the extent that it does not contribute to the objective, despite the fact that the soldier himself does not care about these things.[106]

The considerations I have appealed to in this section are meant to motivate the idea that directives interactions involve a special form of influence, which is not best described, from an internal, agential perspective, in terms of the giving of reasons. Proponents of reason-giving accounts often appear to partially acknowledge this, since it is widely recognized that the director's will is meant to replace the addressee's will as a guide to action (e.g. Hart p. 100). I think that philosophers have been reluctant to abandon the reason-giving paradigm because it has often been tacitly assumed that rational agents are closed intentional systems. Agents are taken to be discrete units with impermeable boundaries, which clearly demarcate them from the world (including other agents). The rational agent acts by first figuring out what she has

[105] For a discussion of the contexts in which purposes external to an acting agent can and cannot be attributed to that agent's action, see Hanser (1998). Hanser argues that a commander's purpose is attributable to an addressee's action when the commander's purpose (non-deviantly) explains why the addressee's action was performed (pp. 390-391). Hanser's view complements the view I am proferring here.
[106] His action might also have the purpose of making money insofar as it is the product of a disposition that exists for that purpose.

reason to do given what she wants and what the world is like and then shaping the world in accordance with her figuring; there is no other (rational) way. Under this assumption, the reason-giving paradigm is the only plausible paradigm for understanding what happens when a rational agent responds to directives, since directives are just another part of the world to which the agent must react. But when one pays close attention to what it is really like to be a socially-embedded agent, the assumption begins to look both overly simplistic and overly intellectual. The assumption looks overly simplistic in that it fails to account for the fact that agents sometimes seem to act rationally and for reasons--in other words, in the full blown sense--without processing the reasons for which they act. On the other hand, the assumption looks overly intellectual because it *requires* agents to recognize the reasons for which they act. Take the case of a child who is directed by his mother to look both ways before crossing the street. The child may not understand that he is doing so for his own safety. But, intuitively he is doing it for that reason because that is the point of the action that his mother recognized in telling him to do it.[107]

I think we should reject the idea that agents are closed intentional systems. Agents have porous boundaries, which enable them to share intentions and reasons with one another. This means that sometimes one has to look beyond the agent herself to delimit her activities. Her plans and projects will sometimes find expression in the actions of others. And, in turn, her own actions may embody someone else's intentions and reasons. The frequency with which any given agent directs and is directed by others will depend on various factors, including one's age, culture, political environment, socioeconomic status, profession, gregariousness, social

[107] Nevermind the fact that a 36-month-old child may not have the conceptual capacities required to recognize a directive-reason, despite the fact that 36-month-old children do respond to directives regularly. For more on the role of directives in the lives of young children, see Chapter V.

connectedness, individual dispositions, and so on. But I think that most socially embedded agents participate in directive interactions a lot more often than they realize. In some big ways and in countless small ones, socially embedded agents are inexorably intertwined with one another.

Let me sum things up and then make a few clarifying remarks. Human beings direct one another all the time. Philosophers have tended to think that directives operate by giving addressees reasons. I have denied this. If we take a close look at how directives seem to influence practical thought, it becomes clear that directive interactions characteristically involve a kind of interpersonal intention transmission. This insight led us to recognize that socially embedded agents are not closed intentional systems. Rather, socially embedded agents inhabit networks of agents that are interconnected by shared intentions, plans, and projects.

At this point, of few clarificatory remarks are in order. First, there is a difference between letting someone else decide what you will do and blindly doing what someone tells you to do. Directives only work as mechanisms of interpersonal intention transmission when addressees recognize, at least tacitly, that issuers are in a position to settle practical questions on their behalf. We attribute this standing to one another in a wide variety of situations, as evinced by our disposition to comply automatically with all sorts of directives. Normally, however, this power is recognized to be limited to some restricted domain. Directives falling outside of that domain or having some exceptional characteristic will not be recognized as authoritative by the addressee. For instance, although an employee recognizes that his employer can decide what he will do at work, the employee will not recognize as authoritative any directive that applies to his personal life or that requires him to do something ridiculous, like jump off a cliff. This is not because, when he receives a directive from his employer, the employee must deliberate about whether to comply. Rather, it is because he does not recognize that his employer is in a position to make those sorts of decisions for him.

Additionally, addressees will typically be disposed to override issuers' intentions if circumstances change in certain ways (just as one is disposed to override one's own previous decisions in light of new information). For instance, if an employee encounters someone needing immediate assistance while delivering a package for his employer, he will, if he is acting rightly, abandon his employer's task and help the person in need. This is compatible with understanding directives as intentions rather than reasons because acting on a transmitted intention does not entail obsequiousness. All it entails is that the addressee is defeasibly committed to being part of a process wherein the issuer's agency is temporarily extended beyond its normal range.

Second, there will be non-paradigmatic cases wherein someone directs an addressee to φ without the addressee's recognizing the issuer's authority to do so. Typically, in such cases no intention will be transmitted even if the addressee ends up φing. However, as already mentioned, one can imagine an addressee who deliberates about whether to comply after the directive has been issued in a way that is not sensitive to the content of the directive (except insofar as it bears on determining the domain of the directive) but rather concerns only the issuer's standing to direct. If the addressee φs on the basis of deliberation of this kind, then perhaps it is right to say that the addressee has decided to let the issuer decide what he shall do and has taken up his intention after all.

Third, not all directives express intentions. Philosophers have noted that directors sometimes direct addressees when they do not intend or desire the addressee to comply. For instance, a sadistic drill sergeant might command a recruit to perform onerous tasks while believing and intending that the recruit will fail to perform the tasks (Hart 1990 p. 95). Because one's intentions must be consistent with one's other intentional states, the sergeant cannot plausibly be said to have the intention that her directive purports to express. In general, aberrant directive interactions like this one will not involve interpersonal intention transmission. Still, if the addressee believes that the issuer sincerely intends for him to act as directed, the addressee may form without deliberation an intention that corresponds to the intention purportedly expressed by the directive. When this occurs, the teleological structure of the addressee's action will not reflect the issuer's practical reasoning; at most, it will reflect the addressee's reasons for being disposed to comply. Since I am interested in how directives characteristically influence addressees, I will ignore aberrant cases of this kind going forward.[108]

Fourth, interpersonal intention transmission is distinct from the phenomenon of proxy agency. Proxy agency occurs when one agent, an executor, does something that counts as a distinct individual or group's doing something for certain purposes. For example, a lawyer's filing of bankruptcy papers on behalf of a client counts as the client's filing bankruptcy for legal and financial purposes. Kirk Ludwig has recently noted that proxy agency can only occur in social contexts wherein certain people and actions, given certain conditions, are collectively recognized as having statuses that enable them to play roles as proxy actors/actions in certain social transactions (2014, 2018).[109] An executor is a proxy agent in virtue of his being authorized to play that role against a social background where such a role, when appropriately authorized, is collectively recognized. And, similarly, proxy actions play the role they do in

[108] My thinking about this issue was heavily influenced by Matthew Hanser's nuanced discussion of the same topic (2016 pp. 118-119).
[109] Ludwig here builds upon John Searle's notion of a status function (Searle 1995 pp. 39-43).

virtue of their being produced in a way that is collectively recognized to result in their having a proxy status (Ludwig 2014 pp. 89-92).[110]

Much like directive interactions, proxy action involves a kind of extension of one person's agency through another. The two are distinct phenomena, however, and it is instructive to note how they differ. Proxy agency, much like the value of a dime or the significance of a border wall, is a social artifact since it depends on collective attitudes towards certain phenomena and is always relativized to purposes having to do with social interactions. The features which characterize directive interactions are not in this sense social artifacts. The action of a compliant addressee is attributable to the issuer because the addressee is carrying out the issuer's intentions, not because others view the addressee's action in a certain way. In other words, the mechanics of the process whereby one intention is transmitted to another, and not the process's collectively recognized social status, is what makes directive interactions special. Moreover, a compliant addressee's action is not attributable to the issuer (only) for this or that purpose any more than an ordinary action is attributable to an actor (only) for certain purposes, but rather in an unrelativized, albeit perhaps attenuated, way.[111]

With these clarifications out of the way, I now move to further defense of the intention transmission paradigm. In the next chapter, I aim to further clarify the view and show that this paradigm is both psychologically and naturalistically defensible. I will also examine the normative features of interpersonal intention transmission and argue that it does not inherently involve an unacceptable abnegation of autonomy.

---

[110] According to Ludwig, the particular action of an executor is actually the final stage in a larger collective action which includes as a component the original authorization(s) that imparted the relevant proxy status(es) to the executor and his action. This detail is not crucial for our current discussion.

[111] None of this is to say, of course, that an agent cannot direct another to perform a proxy action. But the concepts have different intensions.

IV. Directives, Practical Thought, and the Boundaries of Agency, Part 2

IV.I. Abstract

In this chapter, I discuss and defend the intention transmission paradigm introduced in Chapter III, according to which issuing a directing is a way of issuing an intention to an addressee, which can be acted upon without independent deliberation on the part of the addressee. First, I show that the intention transmission paradigm is both psychologically and naturalistically defensible. Second, I argue that acting on the intentions of another does not necessarily involve an unacceptable abnegation of one's autonomy. I also suggest that the intention transmission paradigm can illuminate what is wrong with certain forms of domination. Finally, I discuss the implications of the intention transmission paradigm for how we ought to think about socially-embedded agents.

IV.II. The Psychological and Naturalistic Defensibility of the Intention Transmission Paradigm

In this section, I undertake the task of showing that the intention transmission paradigm of directives is both psychologically and naturalistically defensible. Recall that according to this paradigm, to issue a directive is characteristically to issue an intention, which can then be taken up by an addressee and acted upon. In the last chapter, I argued that this paradigm is superior to the commonly endorsed reason-giving paradigm (according to which to issue a directive is to give a reason) because it more accurately reflects the ways in which directives characteristically shape practical thought and it renders intelligible the issuer's perspective on and connection to the addressee and the addressee's action. In this section, I address two important objections to this paradigm: namely, the objection that interpersonal intention transmission is somehow psychologically unrealistic or impossible for human beings, and the objection that the social conception is inconsistent with the event-causal framework of agency.

IV.II.I. The Psychological Possibility of Interpersonal Intention Transmission

Typically, when an agent forms an intention, that intention is a product of the agent's own deliberative processes. An agent's deliberative processes shape the content and normative properties of her intentions, and when an agent acts on her intentions, she acts for the sake of the reasons that led her in her deliberative processes to form the intentions she did. In this way, an agent's intentions are typically closely tied to her deliberations.

The intention transmission paradigm entails that intentions can be transmitted from one person to another in the way that the Practical Intimacy outlines. Practical Intimacy rests on two psychological presuppositions. First, it presupposes that an agent can form an intention for someone else. Second, it presupposes that agents can take up intentions from others and act upon them without engaging in deliberation. Each presupposition may seem unintuitive or strange from a psychological standpoint. My goal in this section is to show that both are defensible ideas.

First let us consider the idea that it is psychologically possible to form an intention with someone else's action as its content. Philosophers have sometimes thought that one can only intend one's own actions.[112] Others have disagreed.[113] One thing that everyone can agree on is that there are some uncontroversial constraints on intending. If one believes that some action

[112] For instance, Baier (1970); Wilson (1989); Stoutland (2002); Thompson (2008 pp. 120-123); Velleman 1996 (Self to Self) p. 70

[113] For instance, Anscombe (2000 § 2); Velleman (1997); Tuomela (2007 p. 84); Ferrero (2013); Bratman (2014 pp. 60-64). My discussion in this section is heavily influenced by Bratman's argument.

or state of affairs is impossible, if one believes that one cannot by one's intention settle the question as to whether some action or state of affairs will occur, then one cannot intend that action or state of affairs. For instance, I cannot intend to walk to Jupiter's moon Europa, because I believe that walking to Europa is impossible. And I cannot intend that there be life on Europa, because I do not believe that my intention can settle the issue as to whether there is life on Europa. Everyone can agree, then, that an agent cannot form an intention whose content includes another agent's action when the intention would violate one of these constraints.

As I have been emphasizing, however, agents sometimes are in a position to settle the question as to what others shall do simply by telling them to do it. This suggests that an intention with someone else's action as its content would not necessarily violate the constraints mentioned above. And I take it this is why our folk psychological practices clearly allow for intentions concerning others' actions. In common parlance, we often speak of an agent's intentions for another agent and of an agent's intention that such-and-such state of affairs, which requires for its realization the action of another, obtain. For instance, a professor might intend for her students to turn in their papers on Monday, and an employer might intend that her customers feel respected by her employees. The disposition to form intentions concerning others' actions arises naturally in creatures like us who have the capacity both to direct others and to reason about what it would be good for others to do. To be sure, an agent who believes that another agent is within her power and who forms an intention for that agent typically must do something, like communicate her intention, if that intention is to be fulfilled. But this does not show that the first agent's intention all along concerned her own action. The object of the first agent's intention is the other's action, and her intention is satisfied (supposing it is satisfied) by the other's action, not by the first agent's taking means to the achievement of her end.[114] I conclude that the first psychological presupposition is unproblematic.

Let us now turn to the second psychological presupposition that underlies the interpersonal intention transmission paradigm, which many will find significantly more dubious than the first. The second presupposition says that it is psychologically possible for an agent to form (and act upon) an intention without considering reasons for or against the intended course of action. An objector might argue that this is not possible for agents like us. Intentions always result from deliberation or similar reason-responsive processes, which process reasons for action. This objection is bolstered by the fact that agents sometimes undoubtedly form intentions on their own by processing reasons in a less than fully conscious way. Consider an example from baseball. When a batter hits a fly ball, outfielders standing in the relevant location immediately start moving towards the area where the ball is going. There is not enough time between when the ball is hit and when the outfielders start moving for the outfielder to think consciously about where to move. Nevertheless, the outfielders are acting on an intention, not reflexively. And in this case it is correct to say that they are acting for reasons that they recognize apply to them, which are determined by the position and trajectory of the ball. What distinguishes such cases is that the processes which enable the agents to respond to reasons are more spontaneous and more closely tied to perception than conscious ratiocination.[115]

---

[114] In many cases, the intender's communicative act functions more like an enabling condition of her intention's realization than a constitutive part of it.

[115] For other examples of reason-responsiveness without ratiocination, see Arpaly (2006) chapter 2. Arpaly argues that agents act for (their own) reasons whenever their actions are causally connected to their reasons in certain non-accidental ways, where this need not involve deliberation.

So, one might object to the interpersonal intention transmission paradigm in this way: although it appears as if addressees characteristically respond to directives by taking up intentions without deliberating or considering reasons, really addressees are merely responding to directive reasons via spontaneous reason-responsive processes like the baseball players. If so, then addressees, like outfielders, are responding to reasons after all, and there is nothing particularly special about directive interactions. And this is what we should expect, since agents cannot form intention without responding to reasons via deliberation (or similar reason-responsive processes).

I have already said why I do not think directive interactions are best thought of in this way. If rational addressees must weigh their reasons (consciously or otherwise) in order to comply, then directives cannot play the role that they appear to. But what can be said to an objector who thinks that all rational action must result from deliberation or spontaneous reason-responsive processes?

One might attempt to appeal to phenomenology. When someone tells me to do something, it seems to me that I sometimes simply do it, or come to intend to do it, without any ado at all. For instance, my spouse and I often go on walks, and when we do my spouse sometimes unilaterally decides which path we will take. This usually occurs without any prior discussion or thought about the matter. And when it does, it seems to me that I just go with it, without considering or responding to any reasons of my own. My spouse's reasons, whatever they are, are reasons enough.

Unfortunately, phenomenological evidence is not likely to be convincing in this dialectical context. An objector could respond by pointing out that my reflections on this matter are influenced by my theoretical commitments. Indeed, an objector might reach different conclusions via introspection for this same reason. And then there is always the possibility that I am deliberating subconsciously, which cannot be ruled out via introspection. So, another tack is required.

The answer may lie in an appeal to folk psychology. Folk psychology seems to allow for the possibility of forming an intention in response to a directive without deliberation of any kind. And if folk psychology allows for this possibility, then we have good reason for thinking it is psychologically possible for human beings to take up intentions without deliberation. In order to illustrate my point, I present a case that, according to our folk psychological practices, should be viewed as involving the formation of an intention without deliberation.

Suppose that one agent, Eumaeus, believes that another agent, Odysseus, can decide what he (Eumaeus) shall do within some domain by simply telling him what to do. There are many reasons why Eumaeus might believe something like this. For instance, Eumaeus might have decided to let Odysseus decide what he shall do; or he might have gradually formed the belief as a result of repeated interactions wherein Odysseus directed Eumaeus to do things. Let us assume that Eumaeus has this belief because he believes that Odysseus is his king and that kings can decide what their subjects shall do. Now suppose that Odysseus tells Eumaeus to bring him his bow at a certain time (an action Eumaeus recognizes to be within the domain of what Odysseus is in a position to tell him to do), and that when the time comes Eumaeus brings Odysseus his bow.

The scenario just described is not farfetched. It involves the most quotidian of interactions--one person telling another person to do something, the other person doing it--and an innocuous belief about what another is in a position to do. And yet this scenario seems to be a case of interpersonal intention transmission. To see why, consider first that it makes no sense to deliberate (consciously or otherwise) about a practical question that one already believes is settled. For part of what it is to deliberate is to view some aspect of one's future as open, to view some issue or question as undecided and yet decidable by one's deliberations.[116] Indeed, the function of deliberation is to settle practical questions that one views as unsettled, to commit oneself to a certain possible future while foreclosing others.[117] So long as we assume that Eumaeus believes that Odysseus can settle practical questions for him by simply telling him what to do, then we must assume that Eumaeus will not normally deliberate (consciously or otherwise) about what Odysseus tells him to do, since doing so would contradict his belief. And yet, when Odysseus tells Eumaeus to do something, Eumaeus will normally view the relevant practical question about what he will do as thereby settled. And to view a practical question about what one shall do as settled is tantamount to having an intention regarding it. So, when Odysseus tells Eumaeus to bring him his bow, Eumaeus comes to possess an intention that does not result from his own deliberations. And if he acts on this intention, then he acts on an intention that originates in Odysseus's deliberations, not his own. Eumaeus's case illustrates that interpersonal intention transmission and, more generally, forming an intention without considering reasons are possible and need not involve any exceptionally mysterious psychological mechanisms.

One might object to this argument by insisting that the only plausible explanation of Eumaeus's action is that when Odysseus tells Eumaeus to bring him his bow, Eumaeus reasons that he should bring Odysseus his bow as a means to satisfying his own personal desire, intention, or goal to do whatever Odysseus tells him to do. Thus, Eumaeus himself settles the

[116] This perspective on one's future in deliberation is a practical postulate, something we must believe to carry on with the activities of agency. Cf. Kant's idea that we must believe for practical reasons that rational beings are free to be guided by reasons (Groundwork 4:448)
[117] Cf. Velleman 1997a p. 34

question about what he will do in the process of reasoning about how best to achieve his desire, intention, or goal. And thus, Eumaeus does not act without considering his reasons for action.

Attributing to Eumaeus a desire, intention, or goal to do whatever Odysseus tells him is not incompatible with the described scenario. But attributing to Eumaeus means-end reasoning of the kind mentioned is, for reasons already stated. To be successful, the objector needs to show that there is some problem in either (i) attributing to Eumaeus the belief that Odysseus can decide what he will do or (ii) the idea that this belief can have practical relevance for Eumaeus. Since there does not seem to be any contradiction—or even any strangeness—in the proposition *Odysseus can decide what I will do*, it is hard to see why someone could not believe it, especially someone who has been thoroughly indoctrinated into a monarchical political system. And since this belief has practical content, it would be a sign of irrationality if the belief had no practical relevance for Eumaeus. Yet it would be incredible if one could not believe something as mundane as this without irrationality. I conclude that the objection fails.

Let me reemphasize something I have already said. Eumaeus's belief does not make him an uncritical robot. If Odysseus tells Eumaeus to do something that Eumaeus recognizes is outlandish or in conflict with Odysseus's or Eumaeus's interests, then Eumaeus may deliberate about how to proceed. This is because his belief (assuming it is not an irrational or pathological one) admits of exceptions. Moreover, Eumaeus may reflect on and deliberate about what to do if relevant changes occur between the time at which Odysseus directs Eumaeus and the time at which the action is to be carried out. More generally, carrying out

Odysseus's decision will require Eumaeus to engage many of his agential capacities, including means-end reasoning. All this is perfectly intelligible from a folk psychological perspective.[118]

I conclude that humans do in fact have the capacity to form intentions without deliberation. And this completes my defense of the two main psychological presuppositions underlying the intention transmission paradigm: first, that agents can form intentions for others; second, that agents can take up intentions from others and act upon them without deliberating. I cannot claim that my arguments in this section are entirely dispositive, of course. But I hope that this discussion has assuaged some powerful prima facie worries.

I now turn to a discussion of the compatibility of the social conception with the event-causal framework of agency.

IV.II.II. A Naturalistic Approach to Interpersonal Intention Transmission

Agents are physical beings that exist in a physical world. Actions are physical events that exist in a closed causal system. Philosophers have long wrestled with how to reconcile these claims with the idea that actions are attributable to agents in a sense that is importantly different than the way in which events are ordinarily attributable to any one of their many causes. The first question for a naturalistic action theory is how to distinguish an agent's doings from the things that happen to an agent given that the former and the latter are both components of the great causal web that links all events in our universe, which has its origin outside any of us.

[118] The agential contribution (or potential contribution) of someone like Eumaeus to the process that produces the commanded action will play an important role in my argument for the claim that the event-causal framework of agency is not incompatible with the idea that the commanded behavior is attributable to the addressee (as well as the issuer).

The standard response to this problem is to adopt the event-causal framework of agency. According to the event-causal framework, agency is to be explicated primarily in terms of causal relations between an agent's mental states and mental events on the one hand and events like bodily movements on the other. In contemporary action theory, the event-causal framework originates with Donald Davidson (1963; 2013a; 2013b). According to Davidson, actions are bodily movements that are caused in the right way by mental states that rationalize the action under certain descriptions (i.e. they make the action intelligible from the agent's point of view). The event-causal framework has been endorsed, developed, and modified by many (Goldman 1970; Brand 1984; Dretske 1988; Bishop 1989; Mele 2003; Enç 2003, to name a few[119]), and is considered by many to be the only viable theory of agency available.

If the interpersonal intention transmission paradigm were incompatible or in serious tension with the event-causal framework, this would present a problem. My main purpose in this subsection is to show that it is not. There are two ways in which an objector might object to the paradigm on the basis of the event-causal framework. Part of what I have claimed is that in cases of interpersonal intention transmission addressees can act on issuers' intentions and that when this occurs, the resulting action is attributable to the issuer. An objector could argue that this does not fit the event-causal framework, and so is not a respectable idea. I have also claimed that an addressee's action is attributable to the addressee, even when the addressee acts on the issuer's intention. This, too, might constitute fodder for an objection based on the event-causal theory. I will now examine each objection in turn and show why neither claim is in serious tension with the event-causal framework or a naturalistic picture of agency. Along

[119] For detractors, see Anscombe (2000); Chisholm (1964); Ginet (1990)

the way, I will discuss in greater detail the sense in which an addressee's action is attributable to the issuer.

First, I must address a preliminary issue that I have skirted so far in this chapter. How literally should we interpret the claim that addressees act on issuers' intentions? There are two possibilities. On the one hand, we could insist that issuers and addressees literally share a single intention that is stretched between them, so to speak, and is attributable to the issuer. On this understanding, an issuer and an addressee each have token mental states that are proper parts of the shared intention, where these parts may or may not themselves be intentions but function something like them. When an addressee acts on this proper part, the addressee thereby acts on the shared intention. This is an attractive position if one believes that intentions should be individuated by their deliberative source, but it requires one to posit the existence of intentions that are literally present in two different minds. On the other hand, we could maintain that issuers and addressees do not literally share a single token intention. Instead, they have distinct token intentions that have the same content, normative properties, and deliberative source, and when we say that their intentions are the "same," this is what is meant. The intention in the addressee's head is attributable to the issuer insofar as it originates in, and has the content and normative properties it does because of, the issuer's deliberations. This latter paradigm is less weird, and so I will adopt it below. But actually I suspect the issue is largely terminological since the mental states that are present in (and only in) each agent's head will function the same way in either case. So while I adopt the second paradigm and assume that issuers and addressees do not in the strictest sense share an intention, I do not think any of my arguments hang on this. And I think pretty much everything I say below can be formulated in terms of the former proposal.

Now let us consider the first objection based in the event-causal framework. The standard, Davidsonian event-causal line implies that an event is an action of an agent just in case it is a movement of that agent's body that is caused in the right way by certain mental states of the agent, which rationalize the action under a description (1963; 2013a).[120] I have claimed, however, that the addressee's action is normally attributable to the issuer. Yet the causal connection between the deliberations and intentions in the issuer's head and the addressee's action looks quite different than the connection between an agent's mental states and the agent's actions in a typical case of action. So an objector might argue that the addressee's action cannot be attributed to the issuer according to the event-causal framework,.

In responding to this challenge it will be useful to begin by closely comparing the causal connections involved in interpersonal intention transmission with those involved in standard cases. In standard cases, actions are caused by both endogenous and exogenous states and events. Typically, the most salient causes of action are an intention (or similar mental state) of the acting agent and external conditions or events that constitute the action's context. For instance, if I deliberate, form an intention to cross the street when the light turns green, and successfully carry out that intention, then my crossing of the street is caused by both my intention to cross the street (which is itself a causal product of my deliberations) and the light's turning green. We can represent these causes using a diagram:

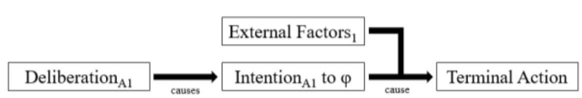

Figure 1

[120] I am ignoring mental actions.

Figure 1 depicts a causal process wherein an agent A1 deliberates, forms an intention, and then performs an action. 'ϕ' in Figure 1 stands for an act-type, and when all goes well Terminal Action will be a token of that act-type, which is caused and rationalized under the relevant description by the intention. Figure 1 is oversimplified in several important respects. For instance, it does not represent any of the causal antecedents to intentions besides deliberation; it does not represent the many other mental states and processes that play a causal role in the production of action; it does not reflect the fact that many external factors cause action via the agent's perceptual faculties; and so on. Nevertheless, Figure 1 captures the basic Davidsonian insight that actions are the causal products of agents' mental states.

Since interpersonal intention transmission is a causal process, it can be represented using a similar but more complex diagram:

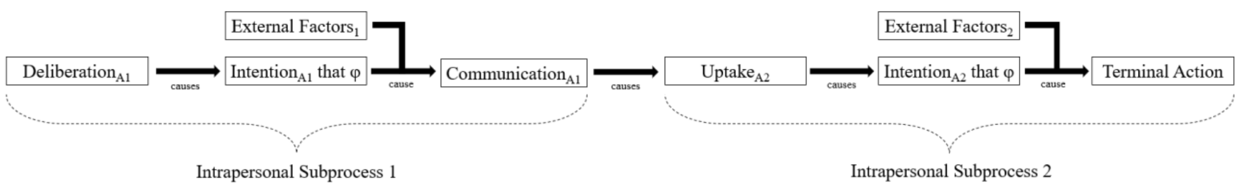

Figure 2

Like Figure 1, in Figure 2 A1's deliberation starts a causal process that terminates in an action. Unlike Figure 1, the causal connection between A1's deliberation and the Terminal Action is interpolated by a communicative action, another agent (A2), and a second intention. A1's intention (Intention$_{A1}$) and A2's intention (Intention$_{A2}$) have the same propositional content. For simplicity's sake, let 'ϕ' stand for a proposition of the form ⌜ A2 ψs ⌝, where 'ψ' stands for an act-type. When all goes well, Terminal Action will be a token of ψ. Intention$_{A1}$ is a product of deliberation, but Intention$_{A2}$ is not (directly) a product of deliberation; it is a product of a communication event involving A1 and A2.

Note that Figure 2 is basically Figure 1 in duplicate. While the overall process represented by Figure 2 is an interpersonal one, it is composed of two subprocesses that are intrapersonal and, except perhaps for the "uptake" box in Intrapersonal Subprocess 2, rather unremarkable. These subprocesses are linked via a communication event. This communication event is constituted by an action of A1 ($\text{Communication}_{A1}$), which terminates Intrapersonal Subprocess 1, and an uptake event (Uptake_{A2}), which begins Intrapersonal Subprocess 2. Since A1 must (at the very least) communicate with A2 if Intention_{A1} is to be fulfilled, Intention_{A1} exerts rational pressure on A1 to communicate with A2 and hence rationalizes $\text{Communication}_{A1}$ under a relevant description. Once uptake occurs, Intention_{A2} exerts rational pressure on A2 to ψ, and thus rationalizes the Terminal Action under a relevant description. In each subprocess, then, the relevant intention plays a role that is analogous to Intention_{A1} in Figure 1.

An example might help make the details of the process depicted in Figure 2 more concrete. Suppose that A1 has agreed to let his teenage son, A2, go on a date tonight but does not want him staying out too late. First, A1 thinks about how best to balance A2's need for independence and desire to have fun with his safety and decides that A2 should return before midnight (Deliberation_{A1}). This decision results in an intention that A2 be home by midnight (Intention_{A1}). When A1 sees A2 (External Factors_1), A1 tells A2 to be home by midnight ($\text{Communication}_{A1}$). A2 agrees to be home at midnight (Uptake_{A2}) and forms an intention to be home before that time (Intention_{A2}). Later, when A2 looks at his watch and sees that it is almost midnight (External Factors_2), he says goodbye to his date and goes home (Terminal Action). A1's deliberation about when A2 should come home started a process that resulted in

A2's arriving home by midnight, an action which was rationalized under that description by the transmitted intention.

We are now in a better position to consider the compatibility of the event-causal framework with the claim that an addressee's action is attributable to the issuer. The relevant sense of attributability is not the one that connects an agent and any given effect of that agent's actions. King Hamlet's death is an event that is attributable to Claudius in this sense, since the king's death is an effect of Claudius's action of pouring poison into the king's ear. While this sense of attributability is important, it is too attenuated to adequately characterize the relation between A1 and Terminal Action. The distinction turns on the role that in each case the agent's intentionality plays in the causal process that connects the agent's actions to the downstream event. The causal process that starts with Claudius's pouring of poison into the king's ear and ends with the king's death is not in any way guided or sustained by Claudius's intentionality--his plans, his reasons, his beliefs, and so on. The existence of this causal process is dependent upon Claudius's intentionality, of course, but Claudius's intentionality does not help explain why the poison, once it was poured into the king's ear, permeated the king's body, curdled his blood, lesioned his skin, and ultimately killed him. On the other hand, the causal process that connects A1 (or, more precisely, $\text{Communication}_{A1}$) and Terminal Action is guided and sustained by A1's intentionality. The content of A1's plan for A2 is interpersonally preserved via the communication event, and this content guides Intrapersonal Subprocess 2 (via A2's agential faculties). Put another way, A1's reasons and plans are integral to an explanation of why $\text{Communication}_{A1}$ causes Terminal Action. For this reason, the connection between A1 and Terminal Action is more intimate than the connection between Claudius and the king's death. While both the king's death and Terminal Action are intended outcomes, A1's but not

Claudius's intentionality extends beyond his action and remains explanatorily relevant to the unfolding of the causal process that results from his action.

The continuous regulation of causal processes by representations (via agential faculties) is one of the characteristic marks of agency.[121] Ballistic actions, wherein the agent's input to a causal process that the agent intends to bring about is limited to the initiation of that process, are relatively rare. Everything from short-lived, simple actions, like the bringing of a cup to one's mouth, to large, complex projects, like the raising of a child or the writing of a book, is filled with obstacles, distractions, fits, and starts, requiring corrections against representations of the goal or plan of action by agential faculties, whether those be subpersonal motor processes, high-level planning faculties, or something in between. The fact that Intrapersonal Subprocess 2 is guided and sustained by A1's intentionality signifies that Intrapersonal Subprocess 2 is an extension of A1's agency, of which Termal Action is a product. Since A1's intentionality regulates the causal process that produces Terminal Action partly by way of another's agential faculties, A1's connection to Terminal Action is not perfectly analogous to A1's connection to A1's own actions, and this shows that the sense in which Terminal Action is attributable to A1 is somewhat unique. Still, the relevant sense of attributability is closer to the sense of attributability that connects an agent to her own actions than the sense that connects her to all the downstream effects of her actions, and for most intents and purposes, including the purposes of attributing responsibility for, rationalizing, and evaluating an action, I think we are justified in treating Terminal Action as if it is A1's action in a wholly ordinary sense (where this does not imply that it is not also A2's action).

[121] The importance of the continuous regulation of actional processes by an agents' intentional faculties for a naturalistic account of agency is emphasized by John Bishop, who appeals to continuous regulation in an attempt to develop a causal deviance-proof analysis of actions (1989).

Part of the upshot of this discussion is that because interpersonal intention transmission is a causal process and because we can explain why and how an addressee's action is attributable to the issuer in causal terms, the claim that the addressee's action is attributable to the issuer is not in conflict with the event-causal framework. Indeed, it would be hard to make sense of this claim without the event-causal framework. My proposal probably requires us to reject the Davidsonian doctrine that an event is an action attributable to an agent only if it is a bodily movement that under some description is a basic action for that agent (2013a), since Terminal Action is probably not best thought of as a basic action for A1 under any description.[122] This is not terribly worrying, for two reasons. First, those who accept my proposal can maintain that Davidson's doctrine is true for all actions not involving interpersonal intention transmission (or: for attributability in the wholly ordinary sense). Second, Davidson's doctrine is not an essential feature of the event-causal framework, as indicated by the fact that many philosophers working directly in the event-causal tradition reject it. For instance, Alvin Goldman (1971) argues that actions are exemplifications of properties at times, and Fred Dretske (1988) argues that actions are causal processes that have intentions (or other mental states or events) and bodily movements (or other events) as components.[123] Hence, the fact that the Social Conception requires us to reject or qualify Davidson's doctrine does not constitute a serious objection to the former.

[122] There might be a way of arguing that Terminal Action is a basic action for A1. Conceivably, there are some cases where Communication$_{A1}$ is not something A1 does in order to bring about Terminal Action. For instance, A2 might be so good at reading A1's mind that A2 can tell what A1 intends by observing his unconscious facial movements. If A1 knows this, and A2 carries out A1's intention without A1's consciously communicating his intention to A2, is Terminal Action a basic action for A1? I find this idea so weird that I do not know what to make of it, but perhaps the answer is yes.

[123] Dretske's process view interestingly leads him to conclude that the actions of others sometimes partly constitute one's own actions. One of Dretske's examples involves a man, J.R., who is divorcing his wife Sue Ellen. Dretske contends that after J.R. has initiated the

At this point, we've gone half of the way to showing that a proponent of the event-causal framework can accept the interpersonal intention transmission paradigm with a good conscience. Now we should consider the second way in which the paradigm might be considered incompatible with the event-causal framework. Recall that in a case of interpersonal intention transmission, the addressee acts upon the issuer's intention (i.e. his action is caused by a mental state that is attributable to the issuer despite being in the addressee's head). On the face of it, this looks to be in tension with the thesis, central to the event-causal framework, that an event is an agent's action only if it is caused by that agent's mental states. If it is the issuer's intention that causes the action, then in what sense is that action the addressee's?

This worry rests on a confusion. Figure 2 can help us resolve it. A2 acts on Intention_{A2}, and thereby produces Terminal Action. A1 is invested in Intention_{A2}; it is A1's in the sense that it originated in and derives its content and normative properties from A1's deliberations. But of course Intention_{A2} is A2's intention in that it is a mental state in A2's head, which is causally and functionally integrated with A2's other mental states and processes. It is within A2's power to dissolve or carry out Intention_{A2}. In short, Intention_{A2} is part of the nexus of intentional states and faculties that constitutes A2 as an agent. So, the objection raised in the last paragraph is a pseudo-problem. Addressee's actions are caused by their own mental states; it is just that these mental states have a special provenance that implicates another agent as well.

While this objection has been easily dispatched, a subtler objection, which does not clearly rest on a confusion, is in the offing. Granted, A2's mental state causes Terminal Action.

process but before the divore has been finalized, J.R. is doing something--*divorcing his wife*--even though in the interim the process is entirely carried forward by others (ibid. pp. 18-19). Basically, for Dretske the entire process represented in Figure 2 (rather than $\text{Communication}_{A1}$ or Terminal Action) should be considered an action that is attributable to A1.

Still, this is only a necessary condition for attributing Terminal Action to A2. We might wonder what about A2's involvement in the process that produces Terminal Action justifies us in describing Terminal Action as an expression of A2's agency. Given that in the event-causal picture there is no uncaused volition or substance causation to which we can point in attributing agency to A2, it may seem as if A2 is a mere conduit in the process depicted in Figure 2. This is a special instance of a general objection facing the event-causal framework. It is sometimes suggested that the cost of producing an analysis of agency wholly in terms of naturalistically respectable and scientifically understandable causal processes is the excision of the agent from that analysis. There is no room, it is said, for an *active* agent to enter the picture so long as the picture is filled only with causally connected physical states and events.[124] And if the agent is not actively involved in the production of an event, then that event cannot really be called an expression of agency--an action--at all.

Many have taken this worry to show that the causal story I have been telling is incomplete. There must be a further element in the analysis of agency, some state, process, or event, that constitutes or signifies the involvement of the agent and sets an expression of agency apart from the product of any other causal process. One popular strategy is to identify the missing element with the operation of a reason-responsive faculty. Berent Enç employs such a strategy. According to Enç, the missing element in the standard event-causal analysis is a computation that involves the weighing of reasons for and against various courses of action--in short, deliberation that is sensitive to reasons and results in decision and intention (2003 ch. 5). According to views like this, the agent's "active" involvement in the causal process that

[124] David Velleman lucidly describes the problem: "In this story, reasons cause an intention, and an intention causes bodily movements, but nobody--that is, no person--*does* anything. Psychological and physiological events take place inside a person, but the person serves merely as the arena for these events: he takes no active part" (1992 p. 461).

produces action is constituted by the process of forming a practical judgment, and the resulting intention and action are attributable to the agent insofar as they are manifestations of this judgment.[125] If the weighing of reasons for and against possible courses of action is a necessary condition for rational action as philosophers like Enç propose, then Terminal Action is not one of A2's actions (at least when Terminal Action does not require any means-end deliberation on the part of A2).[126]

While I ultimately reject Enç's proposal, there is clearly something right in it. The weighing of reasons and the forming of practical judgments are both important elements of agency, and as such a full event-causal analysis of agency should give us an account of these processes in naturalistic terms (which Enç does). But to insist that the agent's "active" contribution to actional processes is wholly or even mainly concentrated in the deliberations that start those processes is to ignore all the ways in which such processes typically involve a continuous practical attentiveness on the part of the agent. As stated above, the continuous regulation of causal processes by representations and agential faculties is a characteristic mark of agency, and most actions, even basic actions, involve regulation of this kind. In general, when an agent intends to do something, that agent is thereby practically committed to a course of action, but that commitment is always susceptible to amendment and rescission should the agent's internal or external circumstances change such that the action no longer makes sense. For instance, a change in an agent's preferences or in the obstacles facing an agent can render an action to which the agent is practically committed all-things-considered undesirable or else

[125] Velleman takes a distinct but similar strategy, arguing that the agent's contribution to action is constituted by the causal contribution of a desire to act for reasons, which, according to Velleman, is constitutive of practical thought itself (1992).
[126] Enç goes so far as to suggest that a mental state is an intention only if it results from (the possessor's) deliberations (2003 p. 171). I argued in the previous subsection that this is a mistake.

render a competing action all-things-considered more desirable. A well-functioning agent will constantly be on the lookout for such changes and, at least in principle, will always be prepared to revise her practical commitments in light of them. I submit that the disposition to rationally revise one's plans in light of changes in the world (or oneself) is just as much a mark of agency as the disposition to rationally change the world (or oneself) in light of one's plans. Agents take an "active" part in actional processes not only by initiating actional processes via rational deliberation, but also by reason-responsively monitoring and adapting such processes.

So, an agent's disposition to rationally revise her practical commitments in response to relevant changes in external or internal circumstances partly constitutes the agent's "active" participation in the causal processes that produce action and thus is part of what justifies us in describing the products of those processes as expressions of her agency. This has implications for the general worry that the agent is somehow lost in the event-causal picture.[127] But more importantly for our purposes, it also explains why we can justifiably describe an addressee's action as an expression of the addressee's agency. Returning to Figure 2, we can see that even when A2 does not need to employ any means-end reasoning to execute Intention_{A2} and even when the time between $\text{Communication}_{A1}$ and Terminal Action is very short, Intrapersonal Subprocess 2 proceeds largely through A2's agential processes and faculties. One consequence of this is that A2 will normally be disposed to rationally revise or abandon Intention_{A2} in response to relevant changes in circumstances. And this is what justifies us in ascribing Terminal Action to A2.

[127] My suggestion can only be part of a fully satisfactory response if it can be shown that the relevant attentiveness is analyzable in event-causal terms. I am fully convinced that it is, but I will not undertake the task here.

This completes my argument for the claim that the event-causal framework is compatible with the interpersonal intention transmission paradigm and in particular with the thesis that the output of a process of interpersonal intention transmission is an action that is attributable to both the issuer and the addressee. In this section as a whole, I have made a first pass case for the claim that the proffered paradigm is both psychologically and naturalistically respectable. I have not addressed all the possible objections in this vein, but I hope to have shown that the paradigm cannot be dismissed out of hand. The reasons for accepting the paradigm derive from the arguments given in the previous chapter and the theoretical work the paradigm can do. Ultimately it must be accepted or rejected on that basis.

In the next section, I turn to a different sort of worry about the interpersonal intention transmission paradigm, namely the claim that interpersonal intention transmission inherently involves domination, subordination, or an infringement of the addressee's autonomy.

IV.III. Autonomy, Directives, and the Interpersonal Intention Transmission Paradigm

In this and the last chapter, I have argued that compliant addressees typically do not deliberate or decide what to do when they are issued a directive. Directives express intentions, and normally a compliant addressee responds to a directive by simply taking up the expressed intention and acting upon it directly. Issuers extends their agency interpersonally by co-opting the actions and agency of addressees for their own purposes; in effect, a compliant addressee is guided by the will of the issuer rather than her own will. One might think that my arguments, if successful, show that there is something inherently objectionable or distasteful about compliance. The ideally rational agent governs herself, and self-governance is incompatible with acting on anyone's intention except one's own. In this section, I address this worry in two ways. First, I argue that one can act on an agent's intention without abnegating one's autonomy, and so there is nothing inherently objectionable about compliance. Second, I argue that abnegating one's autonomy is not always a bad thing.

Let us begin by developing the worry. A good place to start is with Robert Paul Wolff's arguments in *In Defense of Anarchism*. According to Wolff, moral agents are inescapably responsible for their actions because they are free and can reason about what to do. Wolff thinks that agents who are responsible for their actions for this reason have an obligation to be autonomous, to "take responsibility" for their actions whenever possible, where this involves acting on one's own judgment about what to do. Wolff argues that this obligation is in tension with obedience to authority and the duty to obey. Obeying a directive, according to Wolff, involves doing what one is told simply because one is told to do it. An agent who obeys a directive, such as a driver who drives on the right side of the road simply because an official tells her to, acts on another's practical judgment rather than her own and in this way abnegates her autonomy and contravenes her obligation to take responsibility for her actions. The conscientious moral agent may act in accordance with a directive, of course, but only insofar as she judges it all-things-considered right to act as directed, taking into consideration the expectations engendered by the directive, the consequences of her action, and so forth.[128] Hence, Wolff concludes that we do not have an obligation to obey and should not obey (Wolff 1970 pp. 12-17).

Although Wolff's argument is mainly targeted at the legitimacy of political authority, it also applies to the permissibility of complying with any directive of consequence. For

[128] This distinction between obeying and acting in accordance with a directive mirrors Kant's distinction between acting *from* and merely *in accordance with* a duty (Kant 2012 pp. 13-14 (4:397-4:399)).

according to the intention transmission paradigm, addressees do not normally "take responsibility" (in Wolff's technical sense) for their actions when they comply with a directive, since they do not normally act on their own judgment about what to do. Their action appears to be heteronomous, not autonomous. And according to Wolff, we all have an obligation to act autonomously whenever possible.

Wolff takes himself to be developing a Kantian objection to compliance and authority. But Kant himself would not have accepted Wolff's conclusions. In his essay "An answer to the question: What is enlightenment?" (1784/1996), Kant argues that the *public* use of one's reason – by which he means the exercise of one's rational faculties in public discourse – is always permissible and should not be impeded, whereas the *private* use of one's reason – by which he means the exercise of one's rational faculties in deciding for oneself how to act when one is carrying out important social roles – is often impermissible and may be restricted in the interest of a commonwealth. In other words, when one is acting qua taxpayer, soldier, clergyman, *et cetera*, one should sometimes "behave merely passively," since in these scenarios "it is, certainly, impermissible to argue; instead, one must obey" (ibid. p. 18, 8:37). Hence, according to Kant, sometimes it is appropriate for a rational, responsible agent to capitulate to the practical judgment of another (e.g. a legislator).[129] And one can capitulate in good conscience so long as one occupies one's role for good reason and in good conscience.

Kant's discussion of the contexts in which passive obedience is appropriate suggests one way of responding to the worry that compliance with a directive involves an unacceptable abnegation of autonomy. We cannot determine whether an agent is acting autonomously by

[129] Although this discussion is framed in terms of capitulating to another's practical judgment rather than another's intention, this detail is irrelevant, because the effect of capitulating to a practical judgment is the same as the effect of capitulating to another's intention for oneself. In both cases, one acts but not on the basis of one's assessment about what to do.

simply ascertaining whether the intention or judgment on which the agent acts is her own. For agents can autonomously put themselves in positions wherein they are called upon to act on others' intentions or judgments, and, in connection with this, agents can act on more than one intention or judgment at once. Suppose that a man autonomously decides to become a soldier. Part of what it is to be a good soldier is to be disposed to comply with commanders' orders with a certain degree of submissiveness. Hence, to carry out his decision to become a soldier, the man must submit to the will of his commander. If the man's commander (reasonably) orders him to storm a beach, then the man will and should comply with that order by taking up the commander's intention and acting upon it. In storming the beach, the man is acting heteronomously, since he is acting on the commander's intention (judgment) and did not consider the merits of storming the beach himself. But there is also a sense in which the man is acting autonomously. For in complying with the order, the man is acting on his autonomous intention or policy to conduct himself as a soldier should. We might say that the man's action is autonomous relative to one description of the action (*complying with the commander's order*) and heteronomous relative to another (*storming the beach*).

I think the previous example shows that, despite the paradoxical ring, an action can be both autonomous and heteronomous, so long as the agent has autonomously chosen or endorsed her heteronomy. As a first-pass attempt, we might say that an agent's compliance with a directive is autonomous if and only if the agent, in complying, is both (i) carrying out an autonomously chosen course of action which foreseeably involved compliance with a directive like the one being complied with and (ii) able to choose to withdraw from that course of action. These criteria would not count as autonomous the compliance of a soldier who is disposed to comply with his commander's every order because he was forcibly or voluntarily brainwashed. But it would count as autonomous the compliance of a soldier who is disposed to comply because he autonomously endorses the cause for which he fights. And this seems like the right result.

While promising, this first-pass analysis of autonomous heteronomy is too crude. The most glaring problem is that sometimes agents simply find themselves in situations wherein they are disposed to comply with directives, yet intuitively this does not necessarily show that their compliance is non-autonomous. I am riding the elevator to get to the fifth floor of South Hall when suddenly a stranger tells me to hold the elevator door. I comply. Although I realize that such things sometimes occur, it seems a stretch to say that my getting to the fifth floor of South Hall foreseeably involves a directive of any kind. And yet my compliance seems perfectly autonomous. Or consider a teenager who complies with his parent's directive. Typically, teenagers do not choose their parents. And yet, it seems wrong to say that for this reason the teenager's compliant action is non-autonomous. The common element, I believe, is that in each case the agent *would* have autonomously chosen or endorsed a disposition for compliance in these scenarios had the agent considered it beforehand. Facts about what an agent would have endorsed would need to be incorporated into a full account of autonomous heteronomy. Since my purpose here has been merely to show that there is something to the idea of autonomous heteronomy, I will leave the working out of the details for another occasion.[130]

[130] It may be helpful to approach this issue from a slightly different angle. Marina Oshana distinguishes between global and local autonomy. Global autonomy is autonomy relative to one's overall life and values. Local autonomy is autonomy relative to particular times and particular decisions (Oshana 2006). Heteronomy is fully consistent with global autonomy, since an agent's life plan may include her entering into relationships or situations wherein heteronomy is appropriate. And assuming locality comes in degrees, a single action can be heteronomous relative to one degree of locality and autonomous relative to another.

I doubt that Wolff would be happy with this response.[131] His view seems to be that agents have an obligation to avoid heteronomy, period. Still, there is clearly an important moral difference between what I want to call autonomous and non-autonomous heteronomy. All else being equal, it is better to be autonomously heteronomous than non-autonomously heteronomous.

My argument for the claim that heteronomous action can also be autonomous presupposes that we have some grip on the notion of an autonomous choice but is compatible with many ways of cashing that out. Let us briefly survey a few examples. Kant, like Wolff, held that an autonomous choice is one that flows from one's considered judgment about the good. On this view, a heteronomous action would also be autonomous if it resulted from one's judgment that compliance in the relevant situation is good. Harry Frankfurt (1988) argues that autonomy requires that one's lower-order desires and preferences be consistent with one's higher-order desires and preferences concerning one's lower-order desires and preferences. On this view, a heteronomous action would be autonomous if it resulted from a disposition that was produced by desires or preferences that one wants to have. Susan Wolf (1987) argues that an agent is autonomous so long as she has the capacity to track moral reasons. But of course one can comply with a directive while being fully cognizant of the moral reasons that apply to one. Although these accounts of autonomy diverge from one another significantly, they are all consistent in principle with autonomous heteronomy. And we can expect that many other accounts will be similarly consistent.

[131] Evidence that Wolff would not be happy with this response can be found in his discussion of promises to obey. According to Wolff, an agent who promises to obey forfeits his autonomy even, presumably, if the choice to make the promise was autonomous (Wolff 1970 pp. 29-30, 41).

Dialogical accounts of autonomy might seem to present a problem for the possibility of autonomous heteronomy. Roughly, dialogical accounts hold that an important necessary condition of autonomy is that an agent view herself as answerable to others for her practical commitments and actions. In other words, the autonomous agent sees herself as self-responsible, as an appropriate target of criticism and demands for justification (Benson 1994, 2005; Westlund 2003, 2009). As I have emphasized, an agent who acts on another's intention acts for the other's reasons and carries out the other's projects. If I am complying with an order and you ask me why I am doing what I am doing, it would not be infelicitous for me to tell you to ask my commander. There is a sense, then, in which I take my commander rather than myself to be answerable for my action. Still, this by itself does not show that I am non-autonomous. For so long as I take myself to be answerable for my disposition (or decision) for compliance, so long as I would cite my own reasons and commitments if someone challenged that disposition, I am not, as Andrea Westlund puts it, *deeply* deferential. Only a deeply deferential agent is non-autonomous. Thus, even according to dialogical accounts autonomous heteronomy is possible (Westlund 2003 pp. 485-488; 2009 pp. 32-33).

I have just sketched one way of responding to the worry that compliance with a directive always involves an unacceptable abnegation of autonomy. Some heteronomous actions are also autonomous, so acting heteronomously in complying with a directive does not necessarily mean that one is not "taking responsibility" for one's actions. Another way of responding to this worry is to question the idea that abnegating autonomy is always unacceptable.

One approach in this vein is to put pressure on the connection between Wolff's claim that we are responsible for our actions and his claim that we always have to act on our own individual judgments about what is to be done. A major problem with Wolff's argument is that he seems to focus exclusively on individual responsibility while ignoring shared responsibility. Sometimes agents are responsible for an action in virtue of sharing responsibility for that action. Shared responsibility occurs whenever an individual is morally responsible for what other agents or groups do. For instance, a member of an organized political group may share responsibility with others for an act of political expression performed by the group, because the act of political expression is constituted by the member's individual actions in addition to the actions of others. In a situation like this, the group member's (individual) actions gain their moral significance and bear on the member's moral status mainly insofar as they are connected to the activities of the group. The member acts qua group member, and from a moral perspective her actions should be understood in relation to the larger group action of which she is a co-author.

When an agent is committed to participating in a joint enterprise and shares responsibility with others for that enterprise, it is often inappropriate for that agent to unilaterally make and act upon her own (individual) decisions about what to do with respect to that enterprise. Rather, the agent owes it to her fellow group members to act on a *collective* judgment about how to proceed. This is because in such situations, agents cannot be cleanly partitioned into discrete wholes. The agents' actions and moral statuses are intertwined. And the standpoint from which agents in these situations act should reflect this. Hence, sometimes the fact that an agent is responsible for an action shows that the agent *should* abnegate some of their individual autonomy in order to advance the autonomy of a collective to which they belong and in virtue of which they are responsible. For this reason, abnegating autonomy—for instance by following orders issued by a trusted director—may be acceptable or even required of the responsible agent.[132]

Another, more indirect approach to questioning the idea that abnegating autonomy is always unacceptable is to examine to the role of autonomy in a good human life. Clearly, the capacity to direct one's actions according to one's considered judgment is a good thing which should be promoted. Exercising this capacity in relation to matters of fundamental significance to human life and to the arc of one's life as a whole is particularly important. But it does not follow that the capacity should be exercised whenever possible. In fact, there seem to be many goods that are at the very least more difficult to achieve if one insists on always "taking responsibility." Intimate relationships are an important example. Familiars regularly tell one another what to do. This is not only for the reasons outlined in the last chapter (i.e. that doing so facilitates smooth interactions, efficiently utilizes practical knowledge, etc.). To share a life with someone involves forging a joint practical perspective with them, which is constituted by shared interests, values prospects, and so on. Individuals occupying these joint practical perspectives make judgments about what should be done from those perspectives, and these judgments have practical significance for everyone who occupies the perspective. Naturally, familiars who share a practical perspective will be inclined to direct one another in matters relating to these judgments, and familiars will be inclined to accept these directives because and insofar as they align with their shared interests, values, prospects, and so on. Mutual direction stemming from a shared practical perspective is not only a convenient way to get along with someone in the world but may also promote feelings of intimacy, solidarity, and

[132] I explore some of these issues in much greater depth in chapter 5.

love. And this sort of thing would be impossible if one always insisted on "taking responsibility" for one's own actions as Wolff suggests.

What all this suggests is that even if complying with directives sometimes does involve abnegating one's autonomy, this is not unacceptable in all cases. Autonomy is important, but we should not fetishize it to the exclusion of other goods that are important to human life.

In this section, I have argued that compliance with a directive need not involve an unacceptable abnegation of autonomy, since agents can autonomously choose to participate in arrangements wherein they are conduits of the agency of others and since sometimes abnegation of autonomy is acceptable. Of course, sometimes agents do unacceptably forfeit their autonomy in complying. We can even imagine persons whose agency is almost completely enveloped by the agency of others. Thomas Hill Jr. vividly describes such a figure, which he calls the Deferential Wife:

> This is a woman who is utterly devoted to serving her husband. She buys the clothes *he* prefers, invites the guests *he* wants to entertain, and makes love whenever *he* is in the mood. She willingly moves to a new city in order for him to have a more attractive job, counting her own friendships and geographical preferences insignificant by comparison. She loves her husband, but her conduct is not simply an expression of love. She is happy, but she does not subordinate herself as a means to happiness. She does not simply defer to her husband in certain spheres as a trade-off for his deference in other spheres. On the contrary, she tends not to form her own interests, values, and ideals; and, when she does, she counts them as less important than her husband's. (Hill 1973 p. 89)

This form of heteronomy is deeply problematic, and the intention transmission paradigm enables us to vividly see why. The Deferential Wife views herself as a mere instrument of her husband's will. Although the deferential wife does things, she has no significant plans or projects of her own, since her whole life is dedicated to literally carrying out her husband's plans and projects. She is little more than a component in the process whereby her husband shapes the world in his image. And this is incompatible with the wife's dignity as a rational agent with the capacity for self-governance.

So, we can conclude that although complying with directives is not inherently objectionable, too much compliance is not a good thing. I doubt this will strike anyone as particularly surprising.

IV.IV. Conclusion

We have covered a lot of ground. My central purpose in this and the last chapter has been to promote a certain unpopular view about what happens when one agent tells another agent to do something and the other agent does it. Most philosophers have wanted to understand this form of influence in terms of the giving or activating of reasons. I have claimed that this approach will not do. In my view, the relationship between a director and an addressee is much more intimate than that. For an agent who directs another characteristically thereby expresses her intention for the other, and the other who complies with that directive characteristically takes up that intention and produces an action that is in a sense attributable to both. I argued that my view, unlike reason-giving accounts, does justice to the way that directives typically shape the practical thought of the agents who issue and receive them and to the way in which external observers think about the connection between the agents and the resulting action. I then argued that interpersonal intention transmission, as strange as it may sound, is not all that mysterious or objectionable from a psychological, naturalistic, or moral standpoint.

On the surface, this and the last chapter are about directives. But as I have been suggesting throughout, I think that the ideas presented here may have interesting and far-reaching implications for the study of action and agency. Directive interactions pervade our lives. Once one recognizes that many of one's intentions, plans, and projects are carried out by other people and that much of what one does is carry out other peoples' intentions, plans, and projects, it becomes difficult to think of oneself as a discrete entity with clear, impermeable boundaries that correspond perfectly with the boundaries of one's body and mind, as one is tempted to do in solitary introspective moments. This suggests to me that we need to develop a more sophisticated understanding than we now possess of what it means to be a distinct individual that is nevertheless part of a social network of agents that act and think together. If we do, we may discover new avenues for understanding other philosophically important phenomena, such as moral responsibility, communication, joint action,[133] human develop, and much else.[134]

If you, reader, will indulge me, I would like to close by suggesting to you an exercise. Next time you find yourself doing something with someone you know well, like taking a walk, attend closely, if you can, to all the little ways in which your partner directs you hither and thither with subtle and familiar cues. While passing a bench your partner may spontaneously sit, and you may suddenly find yourself sitting as well. Your partner may suggest a detour into a garden to smell some flowers, and you may find yourself heading to that garden to smell those flowers without a thought. You may even find that your partner is just as attuned to you

[133] Cf. Rovane (2014)
[134] Some of these themes are picked up in later chapters of my dissertation.

as you are to them. In moments like these, I think you may find that it is hard to tell where your agency stops and your partner's agency begins.

V. Joint Action Without Robust Theory of Mind

V.I. Abstract

Intuitively, even very young children can act jointly. For instance, a child and her parent can build a simple tower together. According to developmental psychologists, young children develop theory of mind by, among other things, participating in joint actions like this. Yet many leading philosophical accounts of joint action presuppose that participants have a robust theory of mind. In this chapter, I examine two philosophical accounts of joint action designed to circumvent this presupposition, and then I proffer my own novel account of what makes (at least some) interactions between very young children and others joint. I argue that children can take up without deliberation intentions with a joint content that have been transmitted to them by others. In doing so, children can come to share intentions with others, and by acting on these shared intentions they can come to act jointly, all without employing a robust theory of mind.

V.II. Introduction

In recent philosophical literature on joint action, there is growing concern about the compatibility of leading philosophical accounts of joint action with certain research projects in developmental psychology. Developmental psychologists hold that participation in joint action plays an integral role in explaining how children develop a robust theory of mind (RTM). But leading accounts of joint action, like the influential account developed by Michael Bratman (1993; 2014), appeal to shared intentions to explain how joint action occurs, and most analyses of shared intentions presuppose that participants have RTM. This has led to several proposals designed to bridge the gap between philosophical and empirical theory (Tollefsen 2005; Butterfill 2012; Pacherie 2013). Drawing upon recent arguments which purport to show that intentions can be interpersonally transmitted (Roth 2003; 2004; 2014; 2017), I argue that children without RTM can act jointly by acting on a shared intention with a joint content that has been transmitted to them by another. My ultimate goal is to advance our philosophical understanding of the developmentally earliest forms of joint action in which humans can engage.

In section V.III of the essay, I discuss RTM and the mentalizing capabilities of very young children. I then argue that the complex of intentional states which, for Bratman, typically constitutes a shared intention and hence leads to joint action in adults is too psychologically demanding for children. This motivates a search for alternative ways of explicating joint action in very young children. In the second section, I survey two such alternatives due to Stephen Butterfill and Elisabeth Pacherie. In the third and fourth sections, I propose my own novel account. I first describe Abraham Roth's Practical Intimacy thesis, according to which two individuals can share an intention issued from a single deliberative source. I then argue that Practical Intimacy allows those without RTM to share an intention with a joint activity as its content and act jointly.

V.III. The Mentalizing Capacities of Children and Bratmanian Shared Intentions

In this section, I discuss the mentalizing capabilities of very young children and describe RTM. I then examine Bratman's account of shared intentions as it relates to joint action in very young children. I focus on Bratman's account because it is one of the most popular accounts in the literature, but the concerns I direct at (part of) Bratman's account apply to other leading theories for reasons I mention below.[135]

Theory of mind consists in the ability to attribute mental states to agents. There is empirical evidence which suggests that the behavior of very young children in the second year of life (and perhaps earlier) is sometimes informed by an understanding of others' mental states. For instance, very young children appear to form expectations about agents' actions that are sensitive to whether the agent has a true or false belief about the target of the action (Clements and Perner 1994; Onishi and Baillargeon 2005; Baillargeon, et al. 2010). Additionally, two-year-old children can apparently adapt their requests to a communicative partner's relevant knowledge states (O'Neill 1996). And children even sometimes interpret moving objects that look nothing like familiar agents as agents and form expectations about the movement of those objects by attributing intentional states to them (Surian and Geraci 2012; Adam, et al. 2017). All this suggests that very young children have some understanding of agency, beliefs, and goals as well as some ability to attribute these to others.

While very young children have some understanding of mental states, this understanding is at the outset rudimentary, inflexible, and likely implicit (Tager-Flusberg and Sullivan 2000; Apperly and Butterfill 2009).[136] More advanced or explicit theory of mind, as

[135] Much of what I say in this section has been rehearsed in other places (Tollefsen 2005, pp. 77-84; Butterfill, 2012 pp. 24-32; Pacherie 2013, pp. 1823-1826).

[136] It is not entirely clear how we should think about the difference between implicit and explicit theory of mind. Psychologists sometimes note that mental state representations in very young children seem to influence involuntary behavior (e.g. gazes) more readily than voluntary behavior (e.g. verbal responses), and this is perhaps crucial. The concept of access-consciousness might be helpful (Block 1995). One might say that an agent's representations of mental states are explicit to the extent that they are access-conscious. If this is right, then mental state representations start to become explicit at a young age, as indicated, for instance, by the

measured by, for instance, the capacity to represent higher-order mental states and to reason directly about others' mental states, develops later in childhood and depends partly on conceptual and linguistic advancements (Osterhaus et al. 2016; San Juan and Astington 2012).[137] This developmental progression highlights the fact that possession of theory of mind is not a binary trait. An agent's theory of mind can be more or less robust. Let us say, roughly, that RTM is characterized by the ability to represent and reason about mental states – including higher-order mental states – of oneself and others, in addition to an understanding that others' mental states may differ from one's own and that others' mental states may go unsatisfied.[138] Children are not born with RTM; rather, it develops gradually throughout early childhood and beyond.

Now that we have briefly discussed RTM and the mentalizing capabilities of very young children, let us turn to Bratman's influential account of joint action. According to Bratman, joint actions result from shared intentions. Shared intentions play three functional roles in the practical lives of agents. First, shared intentions coordinate the actions of multiple agents around a joint goal. Second, shared intentions help coordinate agents' plans concerning that goal. Third, shared intentions provide a framework that structures bargaining with respect to how the joint goal will be achieved (Bratman 1993, p. 99). For instance, our shared intention

---

fact that 18- and 24-month-old infants spontaneously intervene in anticipation of action mistakes to correct others (Knudsen and Liszkowski 2012).

[137] Interestingly, theory of mind continues to develop even into late adolescence (Dumontheil et al. 2010; Vetter et al. 2013).

[138] This characterization is a modification of Tollefsen's proposal (2005, p. 81). There are various other ways that RTM could be characterized, many of which are compatible with my argument. All that matters for my purposes is that some distinction between the mentalizing capacities of very young children and the mentalizing capacities of adults can be made, where the former capacities are not sufficient for participation in the sort of shared intentions typically discussed by philosophers.

to build a house will tend to facilitate successful planning about the details of our enterprise, such as who will do which task, and, when we carry out those plans, our shared intention coordinates our activities so that we can get the job done. If there is a conflict about, say, which color to paint the house, our shared intention provides the background against which we can resolve the issue.

The three functional roles I just mentioned characterize shared intentions. According to Bratman, the following complex of intentional states, which I call a Bratmanian Shared Intention, plays these three roles and thus constitutes a shared intention:

You and I share a **Bratmanian Shared Intention** to *J* iff:

1. (a) I intend that we *J* and (b) you intend that we *J*.
2. I intend that we *J* in accordance with and because of (1a), (1b), and meshing subplans of (1a) and (1b); you intend that we *J* in accordance with and because of (1a), (1b), and meshing subplans of (1a) and (1b).
3. (1) and (2) are common knowledge between us. (Ibid. p. 106)

Bratman's proposal has increased in complexity over the years, but this early formulation contains the core of his idea.[139] This version contains only two agents, but the account can be generalized for larger groups.[140] Bratmanian Shared Intentions coordinate a group's activities by exerting rational pressure on participants to form intentions and act in ways that are consistent with the group's *J*ing via the agency of all participants. Clause (2) ensures that participants will not bypass one another's agency in pursuit of their goal (which would be

[139] For a more nuanced articulation, see Bratman 2014, pp. 84-85.

[140] There may be more to the story when we consider the actions of collectives that constitute agents in their own right. For an illuminating discussion of some of the unique features of large, structured groups see French (1979) and List & Pettit (2011).

incompatible with joint action) because it commits participants to *J*ing by way of each participant's intention to *J* and to forming co-realizable (“meshing”) subintentions for achieving *J*.[141] Bratman allows for the possibility that something other than a Bratmanian Shared Intention could fulfill the functional roles of a shared intention (I will suggest another such complex in Section V.VI), and hence Bratmanian Shared Intentions are sufficient but not necessary for a joint action (Bratman 1999, p. 144).

Bratmanian Shared Intentions require participants to represent their own and others’ mental states, specifically intentions in (2) and higher-order beliefs or knowledge-states in (3).[142] Furthermore, participants must have some understanding of what it is for subintentions to mesh (which presumably also requires the ability to reason about mental states), since this is part of the content of the intentions in (2). These representations need not be consciously represented, of course (Bratman 2014, p. 104). Still, Bratmanian Shared Intentions presuppose that participants have RTM, and therefore they are too sophisticated for very young children who lack it.

Why is this a problem? The fact that Bratmanian Shared Intentions are cognitively demanding does not undermine the idea that the complexes function as shared intentions for fully developed humans. But it does mean that humans without RTM cannot share an intention and act jointly via Bratmanian Shared Intentions. Yet children do engage in joint action before they come to possess RTM. On an intuitive level, it is difficult to deny that even very young children play, communicate, eat, and perform various other basic activities together with

[141] For a discussion of meshing, see Bratman 2014, pp. 53-56.
[142] For philosophical accounts of common knowledge, see Lewis 1969, pp. 52-60 and Gilbert 1989, pp. 188-195. Both involve higher-order representations. Because other leading accounts of shared intentions require that participants have common knowledge about something, they are also subject to some of the worries I direct at Bratmanian Shared Intentions.

others. The empirical literature seems to support this commonsensical idea, as there is good evidence that early joint action plays an important developmental role. Some psychologists have explicitly argued that the abilities underlying Bratmanian Shared Intentions arise out of rudimentary joint action with adults in infancy (Brownell 2011). There is also evidence that joint action contributes to the development of, and hence is developmentally prior to, RTM. Tuition is a straightforward example. Children learn about the mind by communicating with caretakers who employ mental state talk and who explicitly teach them about the mind (Clements et al. 2000; Taumoepeau and Ruffman 2006; Taumoepeau and Ruffman 2008; Heyes and Frith 2014). Since communication and tuition typically constitute or occur in the context of joint action, this is a reason for thinking that joint action is developmentally prior to RTM. This implies that Bratmanian Shared Intentions cannot explain all joint action.[143] We need an account of how a child without RTM can act jointly that does not involve Bratmanian Shared Intentions. [144]

Let us now turn to the philosophical literature addressing this problem.

V.IV. Two Attempts at Solving the Problem

There are two general strategies for approaching the problem of joint action without RTM: on the one hand, one can maintain with the likes of Bratman that shared intentions play

---

[143] Additionally, people with cognitive disabilities who have difficulty representing and responding to the mental states of others (e.g. autistic individuals, *see* Baron-Cohen 2001; Frith 2001) can seemingly act jointly, but it is likewise implausible to suggest that their acting jointly can best be explained in terms of Bratmanian Shared Intentions.

[144] Carpenter (2009) argues that Tollefsen underestimates the cognitive capacities of very young children and that even very young children have a limited ability to participate in Bratmanian Shared Intentions. While I disagree with this latter claim, Carpenter and I agree that very young children have some understanding of others' mental states and are motivated to engage in joint action. We also agree that joint action in very young children is "very joint."

a central role in explaining joint actions but offer an account of shared intentions that is less cognitively demanding than Bratmanian Shared Intentions; on the other hand, one can articulate a less demanding form of joint action that does not involve shared intentions at all (Butterfill & Sebanz 2011, p. 145). My argument in this chapter comports with the former strategy, but not everyone goes this route. I will now survey two attempts at accounting for joint action without RTM. Although I raise some worries about these accounts, I do not intend for my discussion to be dispositive. Rather, my main aim is to set the stage for my own proposal. I focus on the work of Butterfill and Pacherie because they have garnered significant attention in recent years. Readers only interested in my positive proposal can skip to Section V.V without losing the thread.

V.IV.I. Butterfill: Joint action without shared intention

Stephen Butterfill (2012) articulates a novel notion of joint action without appealing to shared intentions. This simpler form of joint action, which I call *joint action$_B$*, can be instantiated by children without RTM. A group of agents performs a joint action$_B$ when their actions are driven by a "shared goal." Shared goals, in Butterfill's terminology, are not literally goals. Shared goals are complexes of states identified by their functional role of coordinating multiple agents' actions around some outcome to be achieved as a collective effect of their actions (Ibid., p. 37).[145] Butterfill characterizes shared goals as complexes of (i) relations between actions and outcomes to which those actions are directed and (ii) relevant expectations concerning the participants in the joint action$_B$. Agents share a goal if:

[145] Thus, shared goals fulfill one of the three functions that characterize Bratmanian Shared Intentions. But shared goals do not have the function of coordinating plans or structuring bargaining, and in this way they differ from Bratmanian Shared Intentions (ibid., p. 39)

1. the agents' actions are (or will be) directed towards a single outcome,
2. each agent can identify all other participants,
3. each agent expects other participants to perform actions directed towards the outcome,
4. each agent expects the outcome to occur as a collective effect of their actions (paraphrase, Ibid. 40).

Several agents are driven by a shared goal -- and thereby perform a joint action$_B$ -- when they each perform actions directed towards a single outcome in ways that are rational relative to the associated expectations and when their so acting would tend to coordinate their actions and facilitate the goal's occurrence (Ibid., pp. 43-44).

According to Butterfill, the concept of joint action$_B$ can be deployed to make sense of how children perform joint actions without RTM because shared goals do not presuppose RTM. The only requirements are the capacities to perform, understand in some rudimentary sense, and form expectations about goal-directed actions. As noted in Section V.III, there is evidence that very young children have these capacities.[146] A parent and a child whose actions are directed towards a single outcome, like building a tower, and who form reasonable (possibly implicit) expectations about one another's actions and about the effect of their actions are performing a joint action$_B$ even if the child cannot represent mental states explicitly.

Butterfill succeeds in describing a form of interaction that does not presuppose RTM. As Butterfill himself notes, however, joint action$_B$ is a rather thin notion that does not entirely align with our pre-theoretic intuitions about what counts as joint action. It is compatible with one participant dominating another (Butterfill 2012, p. 45) and with the participants using one

[146] Butterfill cites Gergely et al. (1995); Woodward and Sommerville (2000); Csibra et al. (2003); and Luo & Baillargeon (2005).

another as mere social tools (Pacherie 2013, p. 1821). To see just how thin the notion is, consider an example.

Aaron and Betsy recently divorced acrimoniously. Their adult daughter, Charlotte, is a chess player who is competing in a tournament. Aaron and Betsy each want Charlotte to feel supported, and each intends to attend Charlotte's tournament in order to bring about this outcome. Aaron knows that Betsy has this intention and *vice versa*. However, Aaron (falsely) believes that Betsy believes that Aaron will not attend, and likewise Betsy (falsely) believes that Aaron believes that Betsy will not attend. For Charlotte's sake, neither Aaron nor Betsy will interfere with the other's intention, but the divorcees want to avoid one another and would prefer that the other not attend (each believes that their own attendance would be sufficient to achieve the outcome). Aaron and Betsy both attend the tournament, and each discreetly sits far away from the other in order to avoid attracting the other's attention. Neither divorcee ever notices that the other divorcee notices them. Charlotte notices them both and feels supported.

Aaron and Betsy were performing a joint action$_B$ in supporting Charlotte. Their actions were (separately) directed towards the outcome of Charlotte's feeling supported, each could identify the other, each expected the other to perform actions directed towards the aforementioned outcome, and each expected this outcome to occur as an effect of these actions. Moreover, these expectations functioned to coordinate the divorcees' behavior in bringing about this outcome since, for instance, the expectations prompted each of them to be discreet and thereby avoid a quarrel, which would potentially undermine their goal of supporting Charlotte. Yet it seems to me clear that Aaron and Betsy were not performing a joint action in an intuitive sense. For one, each divorcee believed that the other believed that they would not both be attending. Also, neither divorcee was at all inclined to support the intentions of the other. And, finally, Aaron and Betsy coordinated by simply staying out of one another's way.

Similarly unintuitive cases that also count as joint actions$_B$ can be produced with relative ease. So, many interactions that we would not normally be inclined to describe as joint will count as joint actions$_B$. This does not imply that Butterfill has not identified an important type of interaction, nor does it imply that children do not engage in joint actions$_B$. Still, there seems to be a less attenuated sense in which a parent and a child building a tower are acting together. Moreover, considerations of parsimony suggest that if we can explain joint action in children without appealing to a distinct, unintuitive type of joint action, we should. Below I argue that we can do just that. Children without RTM can act jointly in the full-blown, intuitive sense.

Before moving on, let us survey one more attempt at solving the problem of joint action without RTM.

V.IV.II. Pacherie: Shared intentions through team-reasoning

Elisabeth Pacherie (2011, 2013) proposes a solution to the problem of joint action without RTM by developing a notion of "lite" shared intentions that is less cognitively demanding than Bratmanian Shared Intentions. Pacherie constructs her account out of Michael Bacharach's (2006) theory of team-reasoning. To understand Pacherie's proposal, we need to examine three elements borrowed from Bacharach: group-identification, team-reasoning from a group standpoint, and team-reasoning from an individual standpoint.

Group-identification is a framing phenomenon. A frame is "the set of concepts or predicates an agent uses in thinking about the world" (Bacharach 2006, p. 10). An agent may, at different times, think of herself in terms of her individual personality (e.g. as a unique actor), her relationships (e.g. as a spouse), or the larger collectives to which she belongs (e.g. as a member of the department) (Brewer and Gardner 1996). At any given time, an agent's salient self-concepts constitute her self-frame. Self-frames are context dependent and thus unstable, but some circumstances regularly promote "group-identification," or self-framing in terms of a group to which one belongs (Bacharach 2006, pp. 74-81).

Agents who group-identify reason differently about what to do than agents who do not group-identify, because group-identification involves conceiving of one's own agency as a component part of the group's agency (ibid. 136). Agents who group-identify employ team-reasoning. Team-reasoning involves two steps. In the first step, the agent reasons from the perspective of the group, computing what the group should do (as opposed to what she should do qua individual). This is team-reasoning from a group standpoint. In the second step, the agent locates her own role in the group's action, and reasons that she should perform that action.[147] This is team-reasoning from an individual standpoint.

With these tools in hand, we can examine Pacherie's proposal for shared intention "lite."

P1 and P2 share a **Lite Shared Intention** to A if:

(i) each has a self-conception as a member of the team T, consisting of P1 and P2 (*collective self-framing*)

(i') each believes (i) (*group identification expectation*);

[147] Bacharach thinks that the second step is logically inevitable once the first step is completed, assuming one's contribution is necessary for the group's action, because from A→B we can infer O(A)→O(B), where O(x) stands for 'it ought to be the case that x' (ibid., pp. 136, 152-153 n.16)

> (ii) each reasons that A is the best choice of action for the team (*team-reasoning from a group viewpoint*); and
>
> (iii) each therefore intends to do his part of A (*team-reasoning from an individual viewpoint*). (2013, p. 1833)

Lite Shared Intentions are less cognitively demanding than Bratmanian Shared Intentions. Agents must be able to detect and represent others as agents in either case, but Lite Shared Intentions do not seem to require that agents represent others' intentions, form meta-representations, or form expectations about others' beliefs (ibid., p. 1834; Pacherie 2011, pp. 188-190). And neither group-identification nor team-reasoning seems to require representing the mental states of others, which is why Pacherie argues that children without RTM engage in joint activity by reasoning in this way.

The idea that participation in joint action sometimes has a lot to do with the way one reasons is intriguing (cf. Gold and Sugden 2007) and the application of this idea to children is innovative. There is room to worry about the details. One worry stems from the group identification expectation. (i') does not just involve beliefs about whether others will cooperate, which may or may not imply group-identification,[148] but involves beliefs about how oneself and others self-frame, i.e. how they think about or conceptualize their own agency. Whether (i') requires RTM is unclear, but it surely requires beliefs about team members' mental states. Another worry stems from the cognitive requirements of team-reasoning itself. Bacharach thought that some people might fail to team-reason because of cognitive limitations, implying that he thought it somewhat cognitively demanding (2006, p. 128). Bacharach also

[148] Cf. Bacharach's discussion of the different mechanisms that can lead to cooperative coordination (2006, pp. 122-127).

thought it necessary that the payoff structure for the group (i.e. the group's ranking of its available options) be common knowledge between team members and built common knowledge into the reasoning process at other points (ibid., pp. 123, 151 n.5, 161). Whether Bacharach was correct in thinking that common knowledge is necessary for team-reasoning is unclear. What is clear is that if the cognitive requirements of team-reasoning are too robust, then very young children will not be able to utilize it.

Apart from these questions concerning the details of Pacherie's proposal, I have a more general reservation about the team-reasoning approach. The social interactions in which very young children participate are typically structured and managed entirely by adults. A child's participation seems to be, in a way that will be explored in Section V.VI, initially an extension of the reasoning capacities of more competent partners. Only gradually do children gain the ability to make autonomous contributions to joint action (Brownell 2011). This suggests that it may be a mistake to emphasize the reasoning capacities of children when explicating the developmentally earliest forms of joint action. As will become clear, my positive proposal is partly motivated by a desire to take this feature of early social interactions seriously.

We have briefly surveyed two innovative attempts to explain how joint action can occur without RTM. Although I have raised some *prima facie* worries for these accounts, these proposals demonstrate that the project is not hopeless. I turn now to my positive account.

V.V. Practical Intimacy

In this section, I describe Abraham Roth's Practical Intimacy thesis. In the next section, I argue that Roth's thesis can help us understand how children can act jointly.

Friends of Practical Intimacy believe that there is more than one way to acquire an intention. The standard way of acquiring an intention is by making a decision. I want to break in my new shoes, so I decide to go jogging in order to do so. My decision results in an intention to go jogging, which preserves the psychological and normative force of my decision, and my decision is consummated by my intentionally jogging in order to break in my shoes. In addition to this standard way, Practical Intimacy says that I can sometimes acquire an intention without deliberating, consciously or otherwise, by taking up an intention that has been transmitted to me interpersonally.

> **Practical Intimacy:** It is possible for one individual to take up and to act on an intention formed by another without re-issuing the latter's intention. (Roth 2004, p. 383)

Practical Intimacy presupposes that it is possible both to form an intention whose content includes someone else's action (e.g. I intend that she run) and to communicate such an intention (e.g. by saying "I intend that you run"). Call the person who forms such an intention and successfully communicates it to the distinct person who features into the content of that intention *the issuer*, and call that distinct person *the addressee*. According to Practical Intimacy, the addressee can take up the issuer's intention and act on it as if it were her own, without "re-issuing" it – i.e. without independent deliberation.

The idea of an intentional state's being transmitted from one individual to another in a way that circumvents the usual deliberative channels might seem farfetched or pathological, but there is precedent for it in the philosophy of testimony. Reductionist (Humean) views of testimony hold that your justification to believe what you are told is the justification that results from deliberating about the testimony, the testifier, or the practice of assertion. Anti-reductionist (Reidian) views, however, hold that testimony, like perception, involves a basic non-deliberative species of justification. Some Reidians liken testimony to memory, arguing that testimony preserves beliefs across persons just as semantic memory preserves beliefs across time (Burge 1993; Dummett 1994).[149] Similarly, Practical Intimacy says that intentions can be preserved across persons (Roth 2004, pp. 407-408, n. 54, n. 56).

When an intention is shared via Practical Intimacy, the issuer and addressee come to share an intention with the same content and deliberative source. This has important consequences for the teleological structure of the addressee's action, namely the reasons for which she acts, the descriptions under which her action is intentional, and the relevant standards of success. Ordinarily, the teleological structure of my action is determined by my intention and the practical reasoning behind it. I went running because I wanted to break in my shoes; I was intentionally breaking in my shoes; and if in jogging I had not broken in my shoes, my action would have been a failure. My action would not be for the sake of these reasons nor subject to these standards if my intention were not related to my action in a special way.[150]

Practical Intimacy says that the issuer's intention, and by extension the issuer's practical reasoning, can be related to the addressee's action in just this special way. More precisely, when agent X takes up agent Y's intention and acts on it, the teleological structure of X's action is determined by aspects of Y's practical reasoning in much the same way that the teleological structure of X's action is determined by aspects of X's own practical reasoning when X acts on X's own intentions. In Roth's terminology, X is *entitled* to Y's reasons for intending, even if X is not attending to or aware of Y's reasons (Roth 2014a, pp. 643-645; Roth

[149] Thanks to Arnel Blake Batoon

[150] Specifying this "special way" is notoriously difficult. Deviant causal chains show that causal relations are not sufficient (Chisholm 1966; Davidson 1973).

2017, pp. 82-86).[151] For example, suppose I am birdwatching with a friend. My friend spots a goldfinch in a tree and, so that I can see it, tells me to look that direction. I take up that intention and act upon it. In doing so, I am looking at the tree not (or not just) because my friend told me to. I'm looking at the tree for whatever my friend's reasons are, i.e. so I can see the goldfinch. If I fail to see the goldfinch, then my action is a failure. In this way, I am entitled to my partner's reasons, which determine the teleological structure of my action. There are limitations, of course. If I could not understand the reasons, if I would not or could not endorse them, then I am not entitled to my partner's reasons (Roth 2014, p. 646; 2017, pp. 86-90). But sometimes another can decide for me what I will do in such a way that my actions are an extension of their practical reasoning, not mine.

Practical Intimacy should not be sensationalized. Transmission without deliberation does not entail that the addressee is obsequious or even subordinate. The point is only that sometimes one person can settle a practical question for another while both recognize this. Recognizing that some practical question about what I shall do is settled is tantamount to having an intention concerning it. Hence, when I recognize that an issuer has settled a practical question for me, I form an intention, but I do not deliberate. Nothing about such recognition entails obsequiousness or subordination, because I can recognize another's authority to settle practical questions for me while acknowledging the limits of that authority and while enjoying a reciprocal authority over the other (Roth 2004, pp. 393-395).

At this point, it should be reasonably clear how Practical Intimacy differs from accounts of imperatives according to which, at most, the issuer's imperative generates a reason for the

[151] Notice that in ordinary cases, my own reasons can determine the teleological structure of my actions even if I am not attending to, or have temporarily forgotten, why I am doing what I am doing (Roth 2014a, p. 644; Hanser 2015, pp. 116-117; Roth 2017, p. 78).

addressee in favor of acting in a certain way.[152] I do not have the space here to describe Roth's arguments for Practical Intimacy. Instead, in the next section I argue that Practical Intimacy can help us develop a philosophical account of joint action in children. Children often take up others' intentions and act on them, and when the content of those intentions is a joint action, they act jointly. All this can occur without RTM.

V.VI. Practical Intimacy and Joint Action Without RTM

In this section, I argue that Practical Intimacy provides us with the tools to develop a philosophical account of joint action in children. More specifically, I argue that children, like adults, can plausibly adopt shared intentions via Practical Intimacy (section V.VI.I), and that acting on a shared intention with a joint content that has been transmitted in this way is typically sufficient for joint action (section V.VI.II). Indeed, I will suggest that Practical Intimacy is especially well-suited to this task given that early social interactions are typically structured by adults, not children. I assume throughout that it is possible for adults to act on one another's intentions via Practical Intimacy.

V.VI.I. Practical Intimacy in Children

Children of all ages seem to naturally act on the intentions of others (cf. Roth 2004, p. 385). Impressionable children not only imitate behavior, but they seem to readily adopt the intentions of adults and exchange intentions with other children at an early age. I argue that

[152] For instance, Joseph Raz argues that commands generate protected reasons, which provide addresses with both a first-order reason to ϕ and a second-order reason to exclude from deliberation certain first-order reasons to not-ϕ (Raz 2009a, pp. 17-18). For a hybrid account, according to which commands and requests generate reasons for the addressee to act upon the issuer's communicated intention, see Hanser (2015).

children sometimes adopt shared intentions from adults and that this can be understood in terms of Practical Intimacy.

As noted in Section V.III, children can attribute goals to agents before they possess RTM. For instance, very young children sometimes form expectations about what an agent will do in a way that is sensitive to both the agent's goals and belief states (Clements and Perner 1994; Onishi and Baillargeon 2005; Southgate et al. 2007; Baillargeon et al. 2010). To have a goal just is to have an intention.[153] For instance, to have jogging as a goal of mine just is to have an intention to go jogging, which I am practically committed to pursuing. Insofar as children have some understanding of goals, then, they also have some understanding of intentions. In addition to being able to detect intentions, very young children can, of course, also form intentions themselves. If a child sees me building a tower, she can detect the intention behind my movements and form her own intention to build a tower.

These capabilities are prerequisites of Practical Intimacy. But more is required: very young children must be capable of understanding and then taking up without deliberation intentions communicated to them. In their sociable moods, very young children seamlessly accept invitations and directions from others, eagerly participating in their plans and projects. Friends of Practical Intimacy describe these sorts of interactions as instances of interpersonal intention transmission, and since there is no reason for thinking that an addressee needs to represent what she is doing in order to take up an intention (any more than someone must represent what she is doing to learn from testimony), such a description seems apt. So as not

[153] Someone may disagree: Having a goal is sometimes more like having an aspiration than an intention. For instance, I aspire to be a falconer one day, but I do not have any concrete plans to that effect and this goal does not influence any of my planning. Otiose goals of this kind (assuming they are rightly called "goals") are not relevant to my discussion, so I ignore this complication.

to rely entirely on such general remarks, let us examine an experiment wherein children seem to form shared intentions and act jointly with adults and see whether the interaction can be felicitously described in terms of Practical Intimacy. The point of this close examination is to make plausible using a concrete example the idea that Practical Intimacy can occur in children.

First, a preliminary. Note that adults need not verbalize their intentions to communicate them. For instance, when I worked at a nursery garden, my laconic employer often communicated her intention for me by pointing to a plot and handing me a shovel. This is crucial because if the only way to communicate an intention were to say "I intend ϕ", it would be impossible to communicate an intention to individuals who lack the ability to understand talk about mental states, like very young children.

Warneken, Chen, and Tomasello (2006) conducted an experiment designed to measure the extent to which 18- and 24-month-olds, who lack RTM, can understand and form what the experimenters called "joint goals" with adults. Participants were presented with a task in which a participant and an adult experimenter had to work together by simultaneously performing complementary, coordinated actions. The aim of one of the tasks was to retrieve an object inside a vertically movable cylinder. To succeed, one individual had to push the cylinder up while the other simultaneously retrieved the object from the cylinder. Each participant participated in four trials. Before trial one, experimenters demonstrated the activity to the participant. For each trial, an experimenter chose one of the roles and invited the participant's participation by either nonverbal or verbal cues. During the third and fourth trials, the experimenter stopped cooperating mid-activity and stared at the participant unresponsively for fifteen seconds.

The experimenters measured how often participants successfully coordinated with adults to complete the task and observed how participants reacted to uncooperative adults, with the idea that if participants exhibited high levels of coordination and attempted to reengage uncooperative adults, the best explanation would be that they genuinely formed a joint goal with the adult and understood their role in achieving that goal. In general, both 18- and 24-month-old participants were maximally coordinated when performing the retrieving action, while 24-month-olds were more coordinated than 18-month-olds at the pushing action. Participants of both ages often attempted to reengage uncooperative experimenters (as opposed to trying to solve the task individually or disengaging) and did so at approximately the same rate, although 24-month-olds verbalized more often. The researchers determined that participants did not merely imitate experimenters but genuinely understood and pursued a joint goal with them as evidenced by the high level of coordination and robust attempts at reengagement during uncooperative intervals.

If Warneken, Chen, and Tomasello are right, the children formed joint goals for jointly performing a task with adult experimenters, which explains the children's contributions to the task and their attempts to reengage uncooperative experimenters. And, intuitively, the children were acting jointly with experimenters. These observations, however, leave unanswered what exactly a joint goal is, how it is that children without RTM formed a joint goal, and what features of the interaction made it an instance of joint (as opposed to merely coordinated) action. Practical Intimacy provides plausible answers to these questions.

Note that the children participated in an activity that was structured in advance by adult partners. The children, having seen the activity performed before, apparently understood the activity well enough to realize that it constitutively involved the contributions of two coordinated agents, and they readily played a part in that activity when invited to do so. According to Practical Intimacy, exchanges like this are paradigmatic instances of interpersonal intention transmission. This hypothesis invites us to understand talk of joint goals in terms of shared intentions. While one might wonder whether children without RTM have the capacity all by themselves to form the shared intentions required for coordination in relatively sophisticated joint actions, according to Practical Intimacy they do not have to since, and so long as, they can recognize and take up intentions formed by others (in this case, recognition consists in understanding the activity and the invitation, neither of which presupposes RTM). One might also wonder why we can describe the children's actions as contributions to the joint activity when they cannot describe things in this way. According to Practical Intimacy, this is possible because the children's actions inherited their teleological structure from the experimenter's communicated intention and underlying practical reasoning.[154] And, as I argue in the next subsection, acting on an intention with a joint content that has been transmitted from another is typically sufficient for joint action.

So it looks as if this interaction can plausibly be described in terms of Practical Intimacy, and doing so helps us answer some salient questions we might have had about it. I take it that this interaction resembles many social interactions involving children in nonexperimental settings. This suggests that many quotidian social interactions involving children can be felicitously described in terms of Practical Intimacy. Going forward, I assume this is true.

---

[154] The experimenter has other reasons for participating in the joint actions (including those related to the experiment itself) to which the children are not entitled, because the children could not understand or endorse them.

V.VI.II. Joint action in children

Blake is making cookies. Blake's precocious two-year-old, Lizzy, wanders into the kitchen. Blake beckons her over. "Come help me make cookies!" Lizzy grins and approaches the counter (she loves cookies). Blake hands her an egg and asks her to crack it, mimicking the appropriate motion. Lizzy clumsily cracks the egg and puts it into the bowl. A little later, Blake asks Lizzy to roll the dough into balls, which she does, fastidiously placing them on the cookie-sheet. Blake throws the cookie-sheet into the oven and soon the two are enjoying the fruits of their labor with a glass of cold milk.

This scenario seems to involve two individuals performing a joint action. At every step, Blake regulated the activity: the roles each participant played, the techniques they used, and the overarching strategy within which each role was embedded were all products of Blake's agency. Although Lizzy was not an administrator of their activity, she was an active participant. When Blake invited Lizzy to help, he issued an intention to Lizzie whose content was a joint activity, *making cookies together*, and Lizzy effortlessly took up and acted upon this intention as if it were her own. Lizzy's particular contributions, too, were a product of Blake's intentionality. It may not have occurred to Lizzie that she was cracking the egg to make the cookies chewier or conglobing the dough to give the cookies their shape. Nevertheless, she was doing these things for these reasons because she was entitled to Blake's reasons. If, as Practical Intimacy implies, we can say that both participants acted on a shared intention to make cookies and we can say of their particular actions that they were intentional under descriptions that make reference to the same joint action, then I think we are justified in ascribing joint action to them. Blake and Lizzy were baking cookies together, even though Lizzy's contributions were dependent on Blake.

For some, this will not be convincing. An objector might argue that even though Blake and Lizzy were acting on a shared intention to make cookies, this is not sufficient for joint action. There needs to be a richer intentionality like we find in Bratmanian Shared Intentions, according to which participants explicitly acknowledge the agency of other participants. Without it, Lizzy and Blake are at best performing a proto-joint action that is different in kind from the full-blown joint actions that adults engage in, much like Butterfill suggests.

I do not think this is right. Blake and Lizzy are not engaged in a particularly sophisticated joint action, but it is a full-blown joint action nonetheless. The best reason for thinking this is that Blake and Lizzy are acting on a shared intention that fulfills Bratman's functional criteria outlined in Section V.III. Recall that according to Bratman, shared intentions function to coordinate agents' actions and planning around a goal and structure bargaining with respect to how the goal will be achieved. Blake's and Lizzy's shared intention does these three things. The shared intention is present in both Blake's head and Lizzy's head, guiding their behaviors and coordinating their action. That same shared intention exerts rational pressure on Blake to issue subintentions to himself and Lizzy for accomplishing the means necessary to achieve the goal of baking the cookies, and, in turn, these subintentions coordinate the participants' contributions so that the goal is achieved. And their shared intention structures the bargaining of which Blake and Lizzy are capable. For instance, should Lizzy demand that Blake put a cup of orange juice into the batter, Blake would refuse because doing so would undermine the goal to which they are committed. But should Lizzy demand that sprinkles be put into the batter, Blake might acquiesce provided this is compatible with their overarching shared intention. In this way, the shared intention upon which Blake and Lizzy act fulfills Bratman's functional criteria.[155]

The preceding argument might seem surprising given that Lizzy does not have intentions concerning intentions, nor is there common knowledge. According to Bratman, it is not normally enough that you and I each intend that we *J* in conditions of common knowledge, since each of us might intend that we *J* by bypassing the other's agency; our intentions function to coordinate our activities only if each intends that we *J* via both of our intentions (Bratman 1993, pp. 103-104). While this might be true in cases not involving Practical Intimacy, Blake's and Lizzy's case is special in that Blake and Lizzy do not have distinct intentions that they *J*. Rather, Blake and Lizzy share a single intention with a joint content. The satisfaction conditions of any intention include a self-referential component, such that the intention is satisfied only if the intention is the cause of the represented action (Searle 1983, ch. 3).[156] Hence, Blake is committed by his intention to ensuring that he and Lizzy make cookies together via Lizzy's intention that they make cookies together (since it is the same intention). If Blake had an intention concerning the efficacy of Lizzy's intention, it would be redundant, just as it would be redundant for him to have an intention regarding the efficacy of his own intentions. The same goes, *mutatis mutandis*, for Lizzy.

So we now have a philosophical account of what makes at least certain social interactions involving very young children joint: the participants are acting on a shared

[155] It seems as if Bratman would be sympathetic to my argument in this section (Bratman 2014, pp. 104-105).

[156] We probably should not think that this self-referential element is part of the content of the intention in any psychologically demanding sense, however, despite what Searle may have believed. Cf. Burge's (1991) critique of Searle's views concerning self-reference within the intentional content of perceptual representations.

intention with a joint content.[157] And we also have a philosophical account of the mechanism through which the intention is acquired by very young children: interpersonal intention transmission. Practical Intimacy is especially well-suited to describe the rather imbalanced interactions in which children typically participate. Children only gradually become capable of making autonomous contributions to joint action; at the beginning of life, social interactions are structured and managed entirely by adults (Brownell 2011). Philosophical accounts of joint action that require each participant to make more or less symmetrical intentional contributions to joint activities do not comport with what we see in children. As we have seen, sharing an intention via Practical Intimacy involves no such requirement.

In this chapter, I have mainly focused on interactions between children and adults because those occur first developmentally. But Practical Intimacy may also be useful in accounting for joint action between children and their peers. As children gain control over their social environment, they learn to initiate and direct joint activities themselves. This is perhaps best understood in terms of Practical Intimacy since issuing an intention does not seem to require one to represent what one is doing any more than receiving an intention does. Plausibly, then, children learn how to extend their agency through others via Practical Intimacy long before they can represent this process. This is a question for future work.

V.VII. Conclusion

This chapter began with a question: how can we philosophically understand joint action in children without RTM given that leading philosophical accounts of shared intentions presuppose that participants have RTM? This question is significant since intuitively very

[157] My conclusion comports with the spirit of Carpenter's (2009) proposal.

young children can engage in joint action, and, moreover, psychologists tell us that joint action is developmentally important. I reviewed two attempts to answer this question and raised some worries about them. I then introduced Practical Intimacy, according to which one person can act directly on another's intention, and then I argued that young children can act jointly by acting upon an intention transmitted to them by another with a joint activity as its content. My proposal rests upon an empirical premise that children can, without deliberation, take up intentions communicated to them, which I believe to be supported by the evidence. Experiments designed to test this hypothesis could more definitively settle the issue.

It has not been my purpose to account for all instances of joint action involving children or to decisively refute all other extant accounts of what makes these interactions joint. Rather, my aim has been to show the need for further thought on this issue and to introduce a thesis from the literature on action theory that I believe could significantly advance our philosophical understanding in this area. Hopefully this essay will spur further developments in this exciting, cross-disciplinary area of study.

VI. Interpersonal Moral Luck and Normative Entanglement

VI.I. Abstract

In this chapter, I introduce an underdiscussed type of moral luck, which I call *interpersonal moral luck*. Interpersonal moral luck characteristically occurs when the actions of other moral agents, qua morally evaluable actions, affect an agent's moral status in a way that is outside of that agent's capacity to control. I suggest that interpersonal moral luck is common in collective contexts involving shared responsibility and that many philosophers are committed to its existence. I then argue that agents who are susceptible to interpersonal moral luck usually on that basis enjoy special claims against those agents who constitute the source of that luck. This is the phenomenon of *normative entanglement*. I conclude by discussing some of the important ways in which normative entanglement can shape the norms that govern the actions of agents in collective contexts.

VI.II. Introduction

Moral luck occurs when an agent's moral status depends upon factors that are at least partially outside that agent's control. Philosophers have identified many contexts wherein agents seem to be susceptible to moral luck. In this chapter, I discuss moral luck in collective contexts, especially those involving shared responsibility. Shared responsibility occurs when an individual is morally responsible, in the sense of being praiseworthy or blameworthy, for what distinct individuals or groups do. The connection between moral luck and shared responsibility is not discussed much in the literature on either of these topics, and one of my goals is to show that it deserves more attention.

In the section VI.III of the essay, I introduce a type of moral luck, which I call *interpersonal moral luck*, via an example involving shared responsibility. I discuss some of its features and relate it to familiar categories of moral luck. I argue that the distinctive mark of interpersonal moral luck is that it occurs when the actions of others, qua morally evaluable actions, affect one's moral status in a way that is outside of one's capacity to control. In the third section, I argue that agents who are susceptible to interpersonal moral luck sometimes on that basis enjoy special claims against those agents who constitute the source of that luck. This is the phenomenon of *normative entanglement*. I then argue that normative entanglement should affect our thinking about the norms that govern action in collective contexts.

My discussion presupposes the existence of shared responsibility. I will not explicitly argue for this presupposition except insofar as I present cases which, I think, elicit intuitions supporting it. Those who are categorically opposed to shared responsibility can think of my arguments as an exploration of the implications of a false view that many, philosophers and laypeople alike, find plausible.

VI.III. Interpersonal Moral Luck

Consider a case:

> **BANK HEIST:** Robin is recruited to play a role in a bank heist. Robin's task is coordinated in advance since Robin will have no way of communicating with anyone on the day of the heist. On the morning of the heist, Robin will park her car in a prearranged location a few blocks from the bank and then retreat without her car to the countryside so that Robin's coconspirators can escape using Robin's car upon finishing their dirty work in the bank. Everything goes as planned: Robin parks her car, and later Robin's coconspirators rob the bank and use Robin's car to escape.

BANK HEIST involves both a shared intention (to perform a bank heist) and a joint action (heisting a bank). Since Robin intentionally participated in the heist she is morally responsible to some degree for that heist, which is an action that is partly constituted by the actions of other individuals.[158] In other words, Robin shares responsibility for the bank heist.

Given that Robin shares responsibility for the heist, she is subject to moral luck. For if Robin's coconspirators had decided at the last-minute not to perform the heist, Robin would not have been morally responsible for the heist, since there would not have been a heist for which she could be morally responsible (a single parked car does not a heist make). Yet incommunicado Robin cannot control whether her coconspirators decide at the last-minute to go through with the heist. Hence, Robin's moral status -- in particular, her blameworthiness for the heist -- depends upon factors that are partially outside of her control, i.e. her coconspirators' actions.[159] Call this particular way in which Robin is morally (un)lucky *Robin's participatory luck*.[160]

What are the features of Robin's participatory luck? Robin's participatory luck bears similarities to two types of moral luck identified by Thomas Nagel in his initial taxonomy of the subject (1979). *Resultant moral luck* concerns those outcomes of an agent's action that,

[158] Counterfactual accounts of moral responsibility drive a substantial wedge between the degree and scope of moral responsibility (Zimmerman 2002). In this chapter I assume that such views are mistaken and that one's degree of responsibility is closely tied to what one is responsible for. For a comprehensive criticism of counterfactual views, see Hartman (2017) chapter 4.

[159] This is not to deny that Robin would have been blameworthy for something (malicious premeditation, incipient attempt) if the heist had failed to materialize. But she is blameworthy for more and to a greater degree when the heist occurs.

[160] There is, of course, another sense in which Robin's coconspirators' persistence constitutes good luck for Robin since it promotes a goal of hers. This should not worry us. Many cases of moral luck involve such a complexity, e.g. a case of an assassin who fails to kill her target (and thus fails to fulfill her goal but avoids blameworthiness for murder) because a bird intercepts her bullet.

while being in some relevant sense outside of the agent's control, can affect the agent's moral status. *Circumstantial moral luck* concerns an agent's circumstances, insofar as those circumstances, while being outside of the agent's control, can affect the agent's moral status. Let us closely examine Robin's participatory luck in relation to these familiar types of moral luck.

Consider a typical example of resultant moral luck:

> **DRUNK DRIVER:** Gately is driving drunk. A child happens to step in front of Gately's car just as he is driving past, and, because of his drunkenness, Gately cannot react quickly enough to avoid hitting the child. Gately hits and kills the child.[161]

Gately is morally responsible for driving drunk. Gately is also morally responsible for killing the child. DRUNK DRIVER is striking because we take Gately to be much more blameworthy than someone who is morally responsible for driving drunk but not for killing a child, even when this difference in outcomes does not depend upon factors controlled by the blameworthy agents. Gately is morally unlucky because factors outside Gately's control affect his moral status. Notice that, strictly speaking, Gately's moral luck is not explained merely by the fact that Gately could not control whether the child died. If someone else had killed the child, then Gately would not have been morally unlucky, despite the fact that Gately had no control over the child's death. Gately's moral luck is explained by the fact that Gately is not in control of whether his drunk driving results in the child's death (in conjunction with the fact that this relation between his action and the child's death is what makes him responsible). In general, resultant moral luck occurs when an agent is not fully in control of whether certain outcomes,

[161] Both DRUNK DRIVER and the following case (PRISON GUARD) are adapted from Nagel (1979).

which would affect the agent's moral status were they to result from the agent's actions, do in fact result from the agent's actions (Sartorio 2012).

We can imagine many ways in which Robin might be susceptible to moral luck with respect to the outcomes of the heist. Robin has little control over the heist's impact on others, yet the heist's impact can affect Robin's moral status. For instance, while Robin has little control over whether the heist causes intense psychological distress to retirees who store their money in the bank, Robin will be to some degree responsibility for this outcome if and only if the heist does have this outcome.

Robin is not only susceptible to luck with respect to the outcomes of the heist, however, but also with respect to whether she shares responsibility for the heist at all. The issue does not depend on whether the heist is ultimately successful (I assume here that unsuccessful heists are possible). She will be blameworthy for it either way. The issue is that Robin does not fully control whether her action is a constituent part of a larger joint action that is inherently criticizable, and hence she does not fully control whether she is blameworthy for the joint action in virtue of her intentional participation in it.

If Robin's participatory luck turns on whether the heist materializes as an outcome of her action, then BANK HEIST is a special case of resultant luck. There is something strange about this way of characterizing the relationship between Robin's action and the heist, however. Describing something as an outcome of an action (e.g. describing the child's death as an outcome of Gately's drunk driving) usually implies that the thing described is not itself an intentional action. The heist, of course, is an intentional (joint) action, and for this reason describing the heist as an outcome of the individual contributions to it is not felicitous.[162]

[162] At least this seems true for actions of small scale, unstructured collectives. Philosophers have noted that large groups like corporations with decision procedures for determining the

Rather than being related as action is to outcome, the individual participatory actions together constitute the joint action. Saying that the heist is an outcome of Robin's contribution is about as strange as saying that my action of cooking soup is an outcome of my action of getting the pot out of the cabinet. If there is a sense in which the heist is an outcome of Robin's action, I do not think it is a sense relevant to this discussion.[163]

For this reason, I hesitate to categorize Robin's participatory luck as a case of resultant moral luck. Still, Robin's participatory luck in some ways resembles Gately's resultant luck. And this discussion has revealed something important that is worth repeating: strictly speaking, Robin's participatory luck is not explained merely by the fact that she is not in full control of whether the heist occurs, but rather by the fact that she is not in full control of whether her action is a constituent part of the heist (in conjunction with the fact that this constitution relation grounds her moral responsibility). If Robin were fully in control of whether this relation between her action and the heist obtained, then hers would not be a case of moral luck.

Next let us consider circumstantial moral luck. Sometimes circumstantial moral luck is described quite broadly as occurring whenever aspects of an agent's circumstances over which the agent has incomplete control affect the agent's moral status (Hanna 2014 p. 683). Here is a prototypical case:

---

group's actions and attitudes look a lot like agents in their own right (French (1979), List & Pettit (2011)). Perhaps the actions of these groups are best described as an outcome, in the relevant sense, of individual participatory actions.

[163] Certain actions (e.g. murder) can be attributed to an agent only if they produce a certain outcome, and resultant luck with respect to an outcome of this kind is also luck with respect to whether something an agent does that does not entail the outcome (e.g. pulling a trigger) becomes, or comes to constitute, an action that does entail the outcome. This marks a structural similarity between some cases of resultant luck and Robin's participatory luck.

> **PRISON GUARD:** Hans is a career prison guard in Nazi Germany who, although naturally inclined to treat prisoners decently, capitulates to social pressure associated with the regime and mistreats prisoners. Hans has little control over where he lives or his career. But if Hans had immigrated to Argentina before the war, he would have been a virtuous prison guard.

PRISON GUARD describes an individual who is blameworthy for actions that are in an important sense a product of his circumstances, which are largely outside of his control. In general, uncontrolled aspects of one's circumstances often affect the choices, opportunities, or challenges one faces, and these in turn affect what one does and what one is responsible for. When this occurs, one is subject to circumstantial moral luck.

On the face of it, Robin's participatory luck does not look much like Hans' moral luck, which consists in the fact that through no fault of his own he is in a difficult moral situation that leads him to act differently than he would have acted in more favorable circumstances. The sources of Robin's participatory luck are the actions of Robin's coconspirators on the day of the heist, and these are part of Robin's circumstances. But the actions of Robin's coconspirators on the day of the heist do not affect Robin's moral status by affecting the choices, opportunities, or challenges she encounters.[164] They affect it by making it the case that Robin's action is a constituent of a bank heist. Some people reserve the label 'circumstantial moral luck' for cases like PRISON GUARD wherein an agent's circumstances make moral choice more difficult. Robin's participatory luck is not circumstantial in this sense. If, however, circumstantial moral luck occurs whenever an uncontrolled aspect of one's

[164] Of course, Robin's coconspirator's actions before the heist, e.g. their recruitment efforts, might have affected her in this way.

circumstances affects one's moral status as is sometimes suggested, then Robin's participatory luck is circumstantial moral luck.[165]

There is some reason to be cautious at this point. If circumstantial moral luck is defined too broadly, then it will not pick out a distinctive phenomenon. Gately's moral luck in DRUNK DRIVER is a product of his circumstances in the sense that the child's action of stepping into the street is an aspect of Gately's circumstances. Nagel's taxonomy is designed to account for the differences between DRUNK DRIVER and PRISON GUARD, so this is an unattractive result, which suggests that the broad definition of circumstantial moral luck may be unsatisfactory. Ultimately, the issue is terminological and should not worry us overmuch here. What is important for our discussion is how Robin's participatory luck compares to Gately's and Hans'. While it has something in common with each, Robin's participatory luck does not primarily concern the outcome of Robin's action or the choices, opportunities, or challenges she faces. It consists in the fact that the actions of Robin's coconspirators, which are outside of her control, affect whether her action is a constituent of a larger joint action for which, in virtue of this connection, she is blameworthy.

Regardless of how it should be taxonomized, Robin's participatory luck possesses distinctive features that warrant special consideration. One distinctive feature is difficult to describe but relates to how the sources of the moral luck in BANK HEIST compare to the sources

---

[165] In a footnote (1979, p. 34 f. 10), Nagel intimates that we may be able to extend circumstantial luck to cover something like the kind of moral luck I have in mind. He suggests that U.S. citizens with anti-war sentiments were subject to a kind of circumstantial moral luck during the Vietnam War since they had no control over the government's war efforts but were nonetheless implicated in those efforts in virtue of their citizenship. Nagel's discussion in this passage is not entirely clear and the example is controversial. But, in general, I think one can be subject to moral luck in this way. Sometimes an agent shares responsibility for the actions of other members of a group to which the agent belongs simply in virtue of belonging to that group. While it is sometimes tempting to describe this sort of luck as circumstantial moral luck, I am hesitant to do so for reasons I mention in the next paragraph.

of moral luck in the prototypical cases discussed by Nagel and others. The sources of Robin's participatory luck are Robin's coconspirators and their actions. This, by itself, is not distinctive; the source of moral luck in DRUNK DRIVER is, in a sense, the action of the child who steps into the street, and the sources of moral luck in PRISON GUARD are the actions of those who directly and indirectly pressure Hans to act badly. But there is something special about BANK HEIST in that the actions of Robin's coconspirators enter the moral picture qua morally evaluable actions, in much the same way that Robin's own actions do. Indeed, our moral evaluation of Robin as a blameworthy participant in wrongdoing is essentially tied not just to the coconspirators' actions, but to their status as morally evaluable actions. This is not usually true in prototypical cases of moral luck. Even when the sources of the moral luck are actions, they enter the picture qua interloping events. This is reflected in Nagel's discussion of the skeptical import of these sources for our ordinary notions of moral agency and responsibility:

> The problem arises, I believe, because the self which acts and is the object of moral judgment is threatened with dissolution by the absorption of its acts and impulses into the class of events. Moral judgment of a person is judgment not of what happens to him, but of him…The effects of concentrating on the influence of what is not under his control is to make this responsible self seem to disappear, swallowed up by the order of mere events. (Nagel 1979 p. 36)

This passage betrays a conception of moral luck as a product of forces foreign to the domain of moral agency that shape the agent's moral status from the outside. Robin's case is fundamentally different in this respect. A full specification of the salient aspects of her moral position does not bring with it the threat of seeing her agency swallowed up into mere events;

at most, we risk seeing her agency swallowed up into the agency of others. This is important because the moral luck in Robin's situation cannot be explained away by resolving a confusion about what sorts of things are morally evaluable, since moral agents and their intentional actions are incontrovertibly evaluable.

This feature of Robin's participatory luck, then, differentiates it from other types of moral luck. Let us say, somewhat imprecisely, that when the actions of others, qua morally evaluable actions, affect one's moral status in a way that is outside of one's capacity to control, one is subject to *interpersonal moral luck*. A rough-and-ready test for identifying interpersonal moral luck is to ask oneself: can we make sense of the way in which the agent is subject to moral luck without thinking of the actions and agents that are the source of that luck (assuming there are any) *as* morally evaluable actions and agents? Can we make sense of the moral luck if, as Nagel puts it, we view the actions from a standpoint external to agency, as mere "components of the flux of events in the world" (1986 p. 110)? In DRUNK DRIVER, PRISON GUARD, and many cases like them, I think the answer is yes. The child's action in DRUNK DRIVER can be thought of as a mere event in a larger chain of events and the actions of Hans' compatriots in PRISON GUARD can be thought of as fixed psychosocial factors without a serious loss in understanding of the moral luck in either case.[166] This is not true for BANK HEIST. Robin's coconspirators' actions and the joint action as a whole, qua morally evaluable actions, feature essentially into an explanation of Robin's moral responsibility, and hence into an

[166] Perhaps Hans is blameworthy for the German nation's wrongdoing because he is a member of that group (see Jaspers 1947 for an argument for this claim). If so, then he may be subject to interpersonal as well as prototypical circumstantial moral luck. But notice that Hans' would be circumstantially morally unlucky in the same way that he actually is if the pressure he felt to mistreat prisoners were a product of life-like Nazi automata who, in virtue of not being persons, were not morally evaluable. The crux is that he is in a situation wherein moral choice is difficult. It is not an essential feature of his difficulty that the difficulty is brought about by morally evaluable actions.

explanation of Robin's participatory luck. We cannot make sense of what Robin is blameworthy for or why without taking a moral (and agential) perspective on the sources of her moral luck.

There are at least three reasons why interpersonal moral luck deserves special attention. The first I have gestured at already: unlike all the familiar types of moral luck, interpersonal moral luck does not threaten to dissolve our moral conception in the way Nagel feared, and this may have implications for wider discussions about moral luck.

The second reason is that many philosophers writing on interpersonal moral phenomena already seem to be committed to interpersonal moral luck. Although there is room to quarrel about the details, many philosophers accept the general thesis that in collective contexts (paradigmatically, contexts involving joint action), the actions of other others can sometimes affect an agent's moral status.[167] This possibility suggests that collective contexts sometimes bring with them a vulnerability to interpersonal moral luck, since the actions of other participants can sometimes affect one's moral status in a way that is outside of one's control, as BANK HEIST illustrates.[168] Consider a different sort of example. Many philosophers

[167] Two examples of philosophers who are so committed are Larry May and Christopher Kutz. May (1992) argues that agents share responsibility for the wrongdoing of others when, among other things, they causally contribute by act or omission to those wrongdoings or to the communal climate from which those wrongdoings arise. Kutz (2000) argues that the teleological connection between a participant who intentionally participates in a collective action and the shared goals to which that collective action is directed is enough to impute accountability to the participant for the collective's actions, even when the participant is ignorant of the content of the shared goals or of the wrongdoings (pp. 156-159). Both views make room for interpersonal moral luck.

[168] This point should not be overstated. Shared responsibility does not always entail moral luck. Andrew Khoury (2017 pp. 13-18) has argued that it does. Khoury's argument seems to be this: shared responsibility is always grounded in a contingent connection between the object of responsibility (a collective action) and some distinct property of the individual (e.g. a participatory intention), and since the property can exist while the connection fails to obtain, it is always a matter of luck whether the property is related to the object in the appropriate way, and hence whether the individual shares responsibility. The problem with Khoury's argument

would agree that shared responsibility can occur in contexts involving commands. On the one hand, a commander who commands a subordinate to act wrongfully shares some responsibility for that wrongdoing if and only if the subordinate obeys that command. Insofar as the subordinate's actually obeying the command is in some relevant sense outside of the commander's control, the commander is subject to interpersonal moral luck.[169] On the other hand, a subordinate who obeys a malevolent command may be morally responsible for doing so. Thus, a commander's willings can affect a compliant subordinate's moral status, and this may be true even when the commander's malevolence is unknown to the subordinate.[170] These examples expose the close connection between shared responsibility and interpersonal moral luck.

Finally, interpersonal moral luck is of independent interest because its very possibility and presence affects the normative landscape, so to speak, of many interpersonal exchanges. In the next section of the essay I will be occupied with the way in which interpersonal moral luck affects the norms that govern interactions between participants in joint action and the like. The basic idea for which I will argue is this. Often when an agent is susceptible to interpersonal

---

is that it does not follow from the fact that the relation is contingent that the individual is not in control of whether the relation obtains. For instance, suppose that, unlike in BANK HEIST, Robin was not recruited in advance to play a role in the heist. Instead, Robin just happens to see the heist and decides to help the heisters by distracting the bank guards at a crucial moment. In this case Robin shares responsibility for the heist in virtue of her intentional participation in it, but she is in full control over whether she shares responsibility (i.e. of whether the relevant relation between the individual basis of the responsibility attribution and the object obtains). Hence, she is not susceptible to moral luck (in this respect). And hence, shared responsibility does not entail moral luck.

169 This point is due to Gregory Mellema (2016 pp. 83-86). Mellema discusses several ways in which one can be saved from complicity by moral luck, but not all of the ways he mentions should be characterized as interpersonal moral luck.

170 Matthew Hanser's (2015) recent account of commands has this implication. According to Hanser, a subordinate who acts on a command acts for the sake of whatever ends the issuer meant the commanded act to serve, even if the subordinate is unware of what those ends are. Because others' ends are often opaque to us, acting on a command is a morally risky activity.

moral luck, the agent is for that reason defeasibly entitled to demand of the agents who are (or whose actions are) the source of that susceptibility that they avoid acting (or failing to act) in ways that would directly negatively affect the agent's moral status. I call the special relationship marked by this entitlement *normative entanglement*. Normative entanglement, I will argue, is grounded in the interest the agent has in maintaining a certain moral status (or in avoiding a certain moral status and its accompaniments).

VI.IV. Normative Entanglement

To illustrate the phenomenon of normative entanglement and its connection to interpersonal moral luck, let us examine a different sort of case.

> **THIEVING PLUS-ONE:** Marianne is invited to her friend Elinor's house for a party. Marianne wants to bring a plus-one, but the only person available is her acquaintance Willoughby. Marianne knows that Willoughby is a bit of a scoundrel but brings him to the party anyway, and unbeknownst to Marianne, Willoughby steals some of Elinor's electronics.

THIEVING PLUS-ONE exhibits one of the many the dangers inherent in opening one's home to others. Willoughby wrongs Elinor by stealing from her and betraying her hospitality, and for this he is blameworthy. Marianne is also blameworthy. She is blameworthy for putting Elinor at risk by bringing a known scoundrel to the party. But it seems to me that Marianne is more blameworthy than she would have otherwise been had Willoughby not stolen anything. She seems to share some responsibility for Willoughby's action. After all, it would be quite natural for Elinor to blame Marianne somewhat for Willoughby's action and perhaps even to demand that Marianne pay for the stolen property. Given that Marianne is blameworthy because of, or shares some responsibility for, her plus-one's immoral action despite her incomplete control over the situation, Marianne is susceptible to interpersonal moral luck.[171]

Recognizing this puts us in a good position to vindicate a commonsensical way of thinking about THIEVING PLUS-ONE. Suppose Marianne catches Willoughby in the process of pilfering Elinor's electronics. Marianne is entitled to demand—even quite forcefully—that Willoughby cease, and she can do this without the slightest hint of officiousness or impropriety. Now this entitlement is partly explained by Marianne and Elinor's friendship since the norms of friendship empower and require friends to protect one another's interests. But intuitively Marianne's entitlement is much more serious than any other partygoer's entitlement would be (excluding Elinor herself). I contend this is because Willoughby's wrongdoing would directly negatively affect Marianne's moral status, which Marianne has an interest in maintaining, by damaging her moral relationship with Elinor and making her blameworthy for wrongdoing.

Marianne's entitlement to demand that Willoughby conduct himself appropriately at Elinor's party is an example of the phenomenon I call *normative entanglement*. In general, X becomes normatively entangled with Y when X's moral status becomes bound up with Y's activities such that X is entitled to demand, at the very least, that Y avoid gratuitously damaging X's moral status by action or inaction. The entitlements of normative entanglement are

[171] You might have the intuition that in this case Willoughby's action gives Elinor a reason for believing that Marianne is blameworthy but does not actually contribute to her blameworthiness (cf. Richards (1986); Rosebury (1995)). I do not think this is right. In my view, guests have an obligation to make use of hospitality in a way that does not result in harm to the person from whom the hospitality was received. Marianne makes use of Elinor's hospitality in bringing Willoughby as her plus-one, and when Willoughby intentionally steals from Elinor at the party, this constitutes a failure on Marianne's part to make use of Elinor's hospitality without harming her. This failure explains why Marianne shares responsibility for the theft.

grounded in the interest we all have in maintaining an unimpeachable moral status. There is no moral fetishism here. We have an interest in maintaining an unimpeachable moral status because we all have an interest in doing right by others, in maintaining relationships devoid of moral offense, and in faring favorably in others' moral assessments.[172] Marianne, just like all of us, has good reason to want to meet the obligations she owes to her friend and to avoid blameworthiness. Consequently, Marianne is defeasibly entitled to demand that Willoughby not put her in a situation for no good reason wherein she fails to meet her obligations to Marianne and is blameworthy.

One might be tempted to reduce the entitlements associated with normative entanglement to more familiar intragroup norms which exert normative pressure on participants in favor of cooperation in a joint project, like the mutual obligations that, according to Margaret Gilbert (2006), are constitutive of plural subjects. This would be a mistake. First of all, even if we assume that Marianne and Willoughby are participating in a joint action (e.g. attending a party together), it is not clear that the norms internal to their joint action proscribe theft. In any case, interpersonal moral luck as well as normative entanglement can conceivably occur outside the context of joint action.[173] Moreover, normative entanglement can defeasibly entitle an agent to demand that fellow participants refrain from doing their part in a joint action when their doing so would damage the agent's moral status. I think that BANK HEIST is like

[172] Saba Bazargan-Forward (2017) argues in a complementary way that when an agent P2 acts wrongfully in a way that was foreseeably enabled but not condoned by another agent P1, P2 wrongs P1 by making P1 responsible (and hence potentially liable) for the harmful effects of P2's action. According to Bazargan-Forward, P2's action is a setback for P1 because P1 has an interest in maintaining her status as an inviolable being, in avoiding disapprobation, and in not being responsible for a wrongful harm (pp. 119-120).
[173] Nagel's case of the anti-war citizen whose moral status is affected by the U.S.'s war efforts is an example of interpersonal moral luck outside the context of joint action. In cases like this, one is defeasibly entitled to demand that one's fellow group members act in ways that will not negatively affect one's moral status.

this. If Robin could communicate with her coconspirators after she parked the car but before they entered the bank, she would, I think, be defeasibly entitled to demand that they refrain from robbing the bank or, at the very least, from using her car to escape. This sounds a bit strange because the intragroup norms associated with their shared intention prescribe cooperation and thus push in the opposite direction. Perhaps these intragroup norms ultimately defeat Robin's entitlement. The point is that when two or more agents engage in an activity that involves the possibility of interpersonal moral luck, the intragroup landscape may be altered, even if the entitlements associated with normative entanglement are quite circumscribed or ultimately defeated.[174]

The extent to which normative entanglement characterizes our social relations depends upon the pervasiveness of interpersonal moral luck, and if it is never possible for the actions of another to bear on my moral status in the relevant sense, then there will be no normative entanglement.[175] This skeptical position seems overly myopic, however. Many of our most

[174] I have been focusing on the susceptibility of individuals to interpersonal moral luck and normative entanglement. Some of what I say about the susceptibility of individuals might also apply to collectives themselves. Mellema (1997) argues that if individuals are sometimes susceptible to moral luck, then collectives themselves are sometimes susceptible to moral luck. If collectives can have moral statuses in their own right and can enjoy obligations and rights, then perhaps collectives themselves can be susceptible to interpersonal moral luck and become normatively entangled with individuals or other collectives. For instance, plausibly a nation's moral status can be affected by the actions of its soldiers even when those actions are outside of the nation's control. And if so, then the nation, as a collective, may for this reason be in a position to demand that its soldiers avoid acting in ways that would negatively affect its moral status.

[175] Even if there is no interpersonal moral luck and hence no normative entanglement, something similar to normative entanglement may still characterize our social relations. There are many contexts in which the actions of others can affect how one is perceived by others, and in some such contexts it seems for this reason appropriate to make demands of those whose actions can affect one's social status. For example, a faculty member of a philosophy department is arguably defeasibly entitled to demand that her colleagues not negligently embarrass themselves in professional settings, since their embarrassing themselves might reflect poorly on the faculty member's reputation as a member of the department. Another example: A virtuous member of a notorious crime family who has repudiated criminality may

morally important projects, such as child rearing and charity work, are undertaken jointly. To view these projects as consisting in patchworks of discrete moral agents is to ignore the ways in which those who work together for the moral good (or for the moral bad) are in a state of interdependence, not just with respect to their goal, but also with respect to how they should be assessed by others.

This is not to suggest that every instance of interpersonal moral luck involves normative entanglement, for it is possible to subject oneself to interpersonal moral luck in a way that precludes the generation of the relevant defeasible entitlements. Suppose that, without consulting you, I promise a friend that I will get you to perform a supererogatory act, so as to prove to my friend that you are a model of rectitude (a redundant measure, I am sure). Your performance of a supererogatory act is a constitutive condition of my promise's satisfaction, and my promise's being fulfilled (or not fulfilled) does seem to make a difference to my moral status. All else being equal, my moral status is worse if you do not perform the supererogatory act, and I am subject to interpersonal moral luck. Despite this, I think it would be a mistake to say that we are normatively entangled on this basis alone. The possibility of gaining the entitlement to make demands of you in this way, without the input of your agency, seems inconsistent with your autonomy. Still, I think most cases involving interpersonal moral luck will also involve normative entanglement because in most cases the potential for interpersonal moral luck comes about as a product of the agency of both parties.[176]

---

be defeasibly entitled to demand that her less scrupulous kin refrain from wrongdoing, since their committing crimes could give others misleading evidence concerning the virtuous member's blameworthiness. Hence, the phenomenon I am discussing may be of interest even to those who categorically deny the possibility of moral luck. Thanks to an anonymous reviewer for helping me to develop this point.

[176] An exception is the class of cases wherein someone is susceptible to good but not bad interpersonal moral luck. If you cannot possibly damage my moral status, then I am not

The significance of normative entanglement is not merely theoretical. The phenomenon has the potential to impact interpersonal dynamics in important ways. For instance, normative entanglement may facilitate stability in certain joint actions. When we engage in a joint action, there are special difficulties with respect to the stability over time of our shared intentions which arise from the heterogeneity of our individual perspectives, difficulties that do typically not crop-up in the intrapersonal case. As Michael Bratman (2006) points out, if you and I share an intention to paint the house on Friday, there is rational pressure on us favoring the stability of that shared intention, which is supported by instrumental and self-governance considerations. Normative or social-psychological factors that tend to increase the stability of joint projects over time are for this reason desirable.[177] Normative entanglement will sometimes involve entitlements to demand performance in a joint action, thereby generating normative pressure in favor of a joint project, when failure to continue with the project would negatively affect other participants' moral statuses. Hence, normative entanglement will sometimes desirably increase stability.

Besides playing a stabilizing role in joint action, the claims associated with normative entanglement may help explain why people are entitled to demand transparency and conscientiousness from their superordinates and subordinates. A picture of moral agency

normatively entangled with you, even if some of your actions have the potential to cast me in a positive moral light. I believe such cases are very rare.

[177] Such factors are much discussed. Some joint action theorists, such as Margaret Gilbert (2006) and Raimo Tuomela (2007), build stabilizing mutual commitments into the very concept of a joint action. Bratman (2006) suggests that obligations grounded in mutual assurances often play a stabilizing role. Seana Shiffrin (2008 pp. 515-516) argues that promises and related commitments enable participants, through exchanges of practical authority, to take a first-person perspective on joint projects, a practice which should facilitate stability. Tollefsen and Gallagher (2016) argue that shared "we-narratives" can similarly play a stabilizing role. The plethora of proposals suggests both that group stability is a real pragmatic concern and that multiple factors can work together in contributing to stability.

which ties individuals together at the intersections of their sociality is one which subjects those agents to interpersonal moral luck, but it is also one that depicts a more collaborative moral world. We may wonder whether a richer conception of this kind is not more desirable insofar as it allows us to make special judgments concerning interpersonal accountability and obligates us to concern ourselves not just with our own moral status, but the moral statuses of those with whom we collaborate.[178] From the perspective of, say, philanthropists working together to relieve global poverty, the normative unity associated with normative entanglement may play a welcomed role in facilitating group cohesion.[179]

More generally, interpersonal moral luck may ground reasons for or against participating in joint actions in the first place. If a certain activity exposes one to interpersonal moral luck, then one puts oneself at moral risk by participating in that activity. The moral risk associated with employment at a shady corporation or in a questionable government administration exerts rational pressure against participation in those endeavors for the conscientious moral agent. At least with respect to cases like these, a willingness to take the potential effect of others' actions on one's own moral status into account in practical reasoning may lead to a better moral world if it contributes to greater conscientiousness about our role in collective harms.[180]

---

[178] Margaret Urban Walker (1991) argues that acknowledging that moral responsibility outruns control enables us to make certain aretaic judgements concerning integrity and grace, and that this is a good thing. Similar considerations may be germane here.

[179] David Owens (2012) argues that we have normative interests in being bound to others in certain ways. These normative interests explain our ability to voluntarily shape the normative landscape in certain ways, e.g. with a promise or consent. Although I am far from sure about this, I am inclined to think that normative entanglement could, in certain circumstances, be a source of intrinsic value for human beings. Just as it might be good for us to be bound by obligations of love and respect in friendship, so it might be good for us to be bound by obligations of mutual concern for one another's moral statuses in certain joint projects.

[180] This is a familiar point. Tracy Isaac (2011) has argued convincingly that a greater understanding of ourselves in relation to the collectives in which we participate can improve

VI.V. Conclusion

It is almost platitudinous to insist that the actions of others can affect an agent's moral status. The child who darts into the street and the Nazis who pressure prison guards to act wrongly affect other agents' moral statuses in a way that is outside of those agents' control. But what has been largely underappreciated by those who think about these issues is the unique way in which others' actions can bear on an agent's moral status in a particular class of situations – those involving interpersonal moral luck. Sometimes, such as in contexts involving shared responsibility, others' actions, qua morally evaluable actions, become relevant to my moral status in a way that I cannot fully control. This should make a difference not only to our moral assessments, but also to how we respond to those whose agency affects our own moral status and to those whom our moral agency affects.

Investigating whether interpersonal moral luck pervades our moral community is important for those who are worried about the extent to which moral responsibility outruns control. The same arguments for and against the possibility of, say, prototypical resultant and circumstantial moral luck may not smoothly apply to interpersonal moral luck, and thus interpersonal luck may pose a new challenge to those who deny the possibility of moral luck altogether. Moreover, the phenomenon is important for anyone interested in elucidating the normative landscape of the moral community. If your moral status can become tied up with my actions and vice-versa, then this may have serious implications for the nature of our

not only our understanding of our obligations (as individuals) in collective contexts, but also of the opportunities available to us for resolving collective harms.

relationship. For these reasons, it is important that philosophers recognize interpersonal moral luck.

VII. Bibliography

Adam, Maurits, et al. "Agency Cues and 11-Month-Olds' and Adults' Anticipation of Action Goals." *Cognitive Development*, vol. 43, 2017, pp. 37–48, doi:10.1016/j.cogdev.2017.02.008.

Alonso, Facundo. "A Dual Aspect Theory of Shared Intention." *Journal of Social Ontology*, vol. 2, no. 2, 2016, pp. 271-302.

Anscombe, G. E. M. *Intention*. Harvard University Press, 2000.

Apperly, Ian A., and Stephen A. Butterfill. "Do Humans Have Two Systems to Track Beliefs and Belief-Like States?" *Psychologial Review*, vol. 116, no. 4, 2009, pp. 953–970, doi:10.1037/a0016923.

Aristotle. *The Basic Works of Aristotle*. Edited by Richard McKeon, Random House, 1941.

Arpaly, Nomy. *Unprincipled Virtue*. Oxford University Press, 2006.

Bacharach, Michael. *Beyond Individual Choice: Teams and Frames in Game Theory*. Edited by Natalie Gold and Robert Sugden, Princeton University Press, 2006.

Baier, Annette. "Act and Intent." *Journal of Philosophy*, vol. 67, no. 19, 2019, pp. 648–658.

Baillargeon, Renée, et al. "False-Belief Understanding in Infants." *Trends in Cognitive Sciences*, vol. 14, no. 3, 2010, pp. 110–118, doi:10.1016/j.tics.2009.12.006.

Baron-Cohen, Simon. "Theory of Mind and Autism: A Review." *International Review of Research in Mental Retardation*, vol. 23, 2001, pp. 169–184.

Bazargan-Forward, Saba. "Accountability and Intervening Agency: An Asymmetry between Upstream and Downstream Actors." *Utilitas*, vol. 29, no. 1, 2017, pp. 110-124.

Benson, Paul. "Free Agency and Self-Worth." *The Journal of Philosophy*, vol. 91, no. 12, 1994, pp. 650–668.

Benson, Paul. "Taking Ownership: Authority and Voice in Autonomous Agency." *Autonomy and the Challenges to Liberalism*, edited by John Christman and Joel Anderson, Cambridge University Press, 2005, pp. 101–126.

Bishop, John. *Natural Agency: An Essay on the Causal Theory of Action*. Cambridge University Press, 1989.

Block, Ned. "On a Confusion about the Function of Consciousness." *Behavioral and Brain Sciences*, vol. 18, 1995, pp. 227–247.

Brand, Myles. *Intending and Acting: Toward a Naturalized Action Theory*. MIT Press, 1984.

Bratman, Michael E. *Intentions, Plans, and Practical Reason*. Harvard University Press, 1987.

Bratman, Michael E. "Shared Intention." *Ethics*, vol. 104, no. 1, 1993, pp. 97–113.

Bratman, Michael E. *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge University Press, 1999.

Bratman, Michael. "Dynamics of Sociality." *Midwest Studies in Philosophy*, vol. 30, no. 1, 2006, pp. 1-15.

Bratman, Michael E. *Shared Agency*. Oxford University Press, 2014.

Brewer, Marilynn B., and Wendi Gardner. "Who Is this 'We'? Levels of Collective Identity and Self Representations." *Journal of Personality and Social Psychology*, vol. 71, no. 1, 1996, pp. 83–93.

Brownell, Celia A. "Early Developments in Joint Action." *Review of Philosophy and Psychology*, vol. 2, no. 2, 2011, pp. 193–211, doi:10.1007/s13164-011-0056-1.

Burge, Tyler. "Individualism and Psychology." *Philosophical Review*, vol. 95, 1986, pp. 3–45.

Burge, Tyler. "Vision and Intentional Content." *John Searle and His Critics*, edited by Ernest LePore and Robert van Gulick, Blackwell, 1991, pp. 195–213.

Burge, Tyler. "Content Preservation." *The Philosophical Review*, vol. 102, no. 4, 1993, pp. 457–488, doi:10.2307/2185680.

Burge, Tyler. "Reason and the First Person." *Cognition Through Understanding*, Oxford University Press, 2013, pp. 383–406.

Butterfill, Stephen. "Joint Action and Development." *Philosophical Quarterly*, vol. 62, no. 246, 2012, pp. 23–47, doi:10.1111/j.1467-9213.2011.00005.x.

Butterfill, Stephen A., and Natalie Sebanz. "Joint Action: What Is Shared?" *Review of Philosophy and Psychology*, vol. 2, no. 2, 2011, pp. 137–146, doi:10.1007/s13164-011-0062-3.

Carpenter, Malinda. "Just How Joint Is Joint Action in Infancy?" *Topics in Cognitive Science*, vol. 1, no. 2, 2009, pp. 380–392, doi:10.1111/j.1756-8765.2009.01026.x.

Chisholm, Roderick. *Human Freedom and the Self*. The Lindley Lecture, University of Kansas, 1964.

Chisholm, Roderick. "Freedom and Action." *Freedom and Determinism*, edited by K. Lehrer, Random House, 1966, pp. 11–44.

Clements, A. Wendy, and Josef Perner. "Implicit Understanding of Belief." *Cognitive Development*, vol. 9, 1994, pp. 377–395.

Clements, Wendy A., et al. "Promoting the Transition from Implicit to Explicit Understanding: A Training Study of False Belief." *Developmental Science*, vol. 3, no. 1, 2000, pp. 81–92.

Csibra, Gergely, et al. "One-Year-Old Infants Use Teleological Representations of Actions Productively." *Cognitive Science*, vol. 27, no. 1, 2003, pp. 111–133, doi:10.1016/S0364-0213(02)00112-X.

Darwall, Stephen. "Authority and Second-Personal Reasons for Acting." *Reasons for Action*, edited by David Sobel and Steven Wall, Cambridge University Press, 2009, pp. 134–154.

Davidson, Donald. "Actions, Reasons, and Causes." *The Journal of Philosophy*, vol. 60, no. 23, 1963, pp. 685–700.

Davidson, Donald. "Freedom To Act." *Essays on Freedom of Action*, edited by Ted Honderich, Routledge & Kegan Paul Ltd, 1973, pp. 137–156.

Davidson, Donald. "Agency." *Essays on Actions and Events*, 2nd ed., Oxford University Press, 2013a, pp. 43–62.

Davidson, Donald. "Intending." *Essays on Actions and Events*, 2nd ed., Oxford University Press, 2013b, pp. 83–102.

Dretske, Fred. *Explaining Behavior: Reasons in a World of Causes*. MIT Press, 1988.

Dummett, Michael. "Testimony and Memory." *Knowing From Words*, edited by Bimal Matilal and Arindam Chakrabarti, Kluwer Academic Publishers, 1994, pp. 251–272.

Dumontheil, Iroise, et al. “Online Usage of Theory of Mind Continues to Develop in Late Adolescence.” *Developmental Science*, vol. 13, no. 2, 2010, pp. 331–338, doi:10.1111/j.1467-7687.2009.00888.x.

Enç, Berent. *How We Act: Causes, Reasons, and Intentions*. Clarendon Press, 2003.

Enoch, David. “Authority and Reason-Giving.” *Philosophy and Phenomenological Research*, vol. 89, no. 2, 2014, pp. 296–332, doi:10.1111/j.1933-1592.2012.00610.x.

Ferrero, Luca. “Conditional Intentions.” *Nous,* vol. 43, no. 4, 2009, pp. 700-741.

Ferrero, Luca. “Can I Only Intend My Own Actions?” *Oxford Studies in Agency and Responsibility Volume 1*, Oxford University Press, 2013, pp. 70–94.

Finnis, J. M. *Natural Law and Natural Rights*. Oxford University Press, 1980.

Frankfurt, Henry. *The Importance of What We Care About*. Cambridge University Press, 1988.

French, Peter. “The Corporation as a Moral Person.” *American Philosophical Quarterly,* vol. 16, no. 3, 1979, pp. 207-215.

Friedman, R. B. “On the Concept of Authority in Political Philosophy.” *Concepts in Social and Political Philosophy*, edited by R. E. Flatham, Macmillan, Ltd., 1973, pp. 56–91.

Frith, Uta. “Mind Blindness and the Brain in Autism.” *Neuron*, vol. 32, 2001, pp. 969–979.

Gergely, György, et al. “Taking the Intentional Stance at 12 Months of Age.” *Cognition*, vol. 56, no. 2, 1995, pp. 165–193, doi:10.1016/0010-0277(95)00661-H.

Gilbert, Margaret. *On Social Facts*. Princeton University Press, 1989.

Gilbert, Margaret. “Walking Together: A Paradigmatic Social Phenomenon.” *Midwest Studies in Philosophy* XV (1990): 1-14.

Gilbert, Margaret. “Obligation and Joint Commitment.” *Sociality and Responsibility: New Essays in Plural Subject Theory*. Oxford: Rowman & Littlefield Publishers, 2000. 50-70.

Gilbert, Margaret. “The Structure of the Social Atom: Joint Commitment as the Foundation of Human Social Behavior.” *Socializing Metaphysics: The Nature of Social Reality*. Ed. Frederick Schmitt. Oxford: Rowman & Littlefield Publishers, 2003. 39-34.

Gilbert, Margaret. *A Theory of Political Obligation: Membership, Commitment, and the Bonds of Society*. Oxford University Press, 2006.

Gilbert, Margaret. “Introduction.” *Joint Commitment: How We Make the Social World*. Oxford: OUP, 2014a. 1-22.

Gilbert, Margaret. “Considerations on Joint Commitment.” *Joint Commitment: How We Make the Social World*. Oxford: OUP, 2014b. 37-57

Gilbert, Margaret. “Three Dogmas About Promising.” *Joint Commitment: How We Make the Social World*. Oxford: OUP, 2014c. 296-323.

Ginet, Carl. *On Action*. Cambridge University Press, 1990.

Gold, Natalie, and Robert Sugden. “Collective Intentions and Team Agency.” *The Journal of Philosophy*, vol. 104, no. 3, 2007, pp. 109–137.

Goldman, Alvin I. *A Theory of Human Action*. Prentice-Hall, Inc., 1970.

Hanna, Nathan. “Moral Luck Defended.” *Nous*, vol. 48, no. 4, 2014, pp. 683-698.

Hanser, M. “Intention and Teleology.” *Mind*, vol. 107, no. 426, 1998, pp. 381–401, doi:10.1093/mind/107.426.381.

Hanser, Matthew. “Doing Another’s Bidding.” *Reasons and Intentions in Law and Practical Agency*, edited by George Pavlakos and Veronica Rodriguez-Blanco, Cambridge University Press, 2015, pp. 95–120.

Hart, H. L. A. “Commands and Authoritative Legal Reasons.” *Authority,* New York University Press, 1990, pp. 92–114.

Hartman, Robert J. *In Defense of Moral Luck: Why Luck Often Affects Praiseworthiness and Blameworthiness*. Routledge, 2017.

Heyes, Cecilia M., and Chris D. Frith. “The Cultural Evolution of Mind Reading.” *Science*, vol. 344, no. 6190, 2014, doi:10.1126/science.1243091.

Hill, Thomas E. Jr. “Servility and Self-Respect.” *The Monist*, vol. 57, no. 1, 1973, pp. 87–104.

Isaacs, Tracy. *Moral Responsibility in Collective Contexts*. Oxford University Press, 2011.

Jaspers, Karl. *The Question of German Guilt*, Translated by E. B. Ashton, The Dial Press, 1947.

Kant, Immanuel. “An Answer to the Question: What Is Enlightenment?” *Practical Philosophy*, edited by Mary J. Gregor and Allen Wood, 1996, pp. 11–22.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. Edited by Mary Gregor and Jens Timmermann, Cambridge University Press, 2012.

Khoury, Andrew C. “Individual and Collective Responsibility.” *Reflections on Ethics and Responsibility: Essays in Honor of Peter A. French*, Edited by Zachary J. Goldberg, Springer International Publishing, 2017, pp. 1-20.

Knudsen, Birgit, and Ulf Liszkowski. “Eighteen- and 24-month-old infants correct others in anticipationof action mistakes.” *Developmental Science*, vol. 15, no. 1, 2012, pp. 113-122.

Korsgaard, Christine M. *The Sources of Normativity*. Cambridge University Press, 1996.

Kutz, Christopher. *Complicity: Ethics and Law for a Collective Age*. Cambridge University Press, 2000.

Lewis, David K. *Convention: A Philosophical Study*. Harvard University Press, 1969.

List, Christian and Philip Pettit. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford University Press, 2011.

Ludwig, Kirk. “Proxy Agency in Collective Action.” *Nous*, vol. 48, no. 1, 2014, pp. 75–105, doi:10.1111/nous.12013.

Ludwig, Kirk. “Proxy Agency in Collective Action.” *The Routledge Handbook of Collective Intentionality*, edited by Marija Jankovic and Kirk Ludwig, Routledge, 2018, pp. 58–67.

Luo, Yuyan, and Renée Baillargeon. “Can a Self-Propelled Box Have a Goal? - Psychological Reasoning in 5-Month-Old Infants.” *Psychological Science*, vol. 16, no. 8, 2005, pp. 601–608, doi:10.1111/j.1467-9280.2005.01582.x.

May, Larry. *Sharing Responsibility*. The University of Chicago Press, 1992.

McMahon, Christopher. *Authority and Democracy: A General Theory of Government and Management*. Princeton University Press, 1994.

McMahon, Christopher. “Shared Agency and Rational Cooperation.” *Nous*, vol. 39, no. 2, 2005, pp. 284-308.

Mele, Alfred R. *Motivation and Agency*. Oxford University Press, 2003.

Mellema, Gregory. “Moral Luck and Collectives.” *Journal of Social Philosophy*, vol. 28, no. 3, 1997, pp. 144-152.
Mellema, Gregory. *Complicity and Moral Accountability*. University of Notre Dame Press, 2016.
Miller, Seamus. *Social Action: A Teleological Account*. New York: Cambridge UP, 2001.
Nagel, Thomas. Moral Luck. *Mortal Questions*. Cambridge University Press, 1979, pp. 24-38.
Nagel, Thomas. *The View from Nowhere*. Oxford University Press, 1986.
O’Neill, Daniela K. “Two-Year-Old Children’s Sensitivity to a Parent’s Knowledge State When Making Requests.” *Child Development*, vol. 67, 1996, pp. 659–677.
Onishi, Kristine H., and Renée Baillargeon. “Do 15-Month-Old Infants Understand False Beliefs ?” *Science*, vol. 308, no. April, 2005, pp. 255–258.
Oshana, Marina. *Personal Autonomy in Society*. Routledge, 2006.
Osterhaus, Christopher, et al. “Scaling of Advanced Theory-of-Mind Tasks.” *Child Development*, vol. 87, no. 6, 2016, pp. 1971–1991, doi:10.1111/cdev.12566.
Owens, David. *Shaping the Normative Landscape*. Oxford University Press, 2012.
Pacherie, Elisabeth. “Framing Joint Action.” *Review of Philosophy and Psychology*, vol. 2, no. 2, 2011, pp. 173–192.
Pacherie, Elisabeth. “Intentional Joint Agency: Shared Intention Lite.” *Synthese*, vol. 190, no. 10, 2013, pp. 1817–1839, doi:10.1007/s11229-013-0263-7.
Raz, Joseph. *The Authority of Law: Essays on Law and Morality*. Clarendon Press, 1979.
Raz, Joseph. *The Morality of Freedom*. Oxford University Press, 1986.
Raz, Joseph. *Practical Reason and Norms*. Oxford University Press, 1999.
Raz, Joseph. “Legitimate Authority.” *The Authority of Law: Essays on Law and Morality*, 2nd ed., Oxford University Press, 2009a, pp. 3–27.
Raz, Joseph. “The Problem of Authority: Revisiting the Service Conception.” *Between Authority and Interpretation: On the Theory of Law and Practical Reason*, Oxford University Press, 2009b, pp. 126–165.
Richards, Norvin. “Luck and Desert.” *Mind*, vol. 95, no. 378, 1986, pp. 198-209.
Rosebury, Brian. “Moral Responsibility and ‘Moral Luck’.” *The Philosophical Review*, vol. 104, no. 4, 1996, pp. 499-524.
Roth, Abraham Sesshu. “Practical Intersubjectivity.” *Socializing Metaphysics: The Nature of Social Reality*, Rowman & Littlefield, 2003, pp. 65–92.
Roth, Abraham Sesshu. “Shared Agency and Contralateral Commitments.” *The Philosophical Review*, vol. 113, no. 3, 2004, pp. 359–410.
Roth, Abraham. “Shared Agency”, *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta, 2011.
Roth, Abraham Sesshu. “Prediction, Authority, and Entitlement in Shared Activity.” *Nous*, vol. 48, no. 4, 2014a, pp. 626–652, doi:10.1111/nous.12011.
Roth, Abraham. “Indispensability, the Discursive Dilemma, and Groups with Minds of Their Own.” *From Individual to Collective Intentionality: New Essays*. Edited by Sara Rachel Chang, Frank Hindriks, and Gerhard Preyer, Oxford University Press, 2014b, pp. 137-162.

Roth, Abraham Sesshu. “Entitlement to Reasons for Action.” *Oxford Studies in Agency and Responsibility Volume 4*, edited by David Shoemaker, Oxford University Press, 2017, pp. 75–92.

Rovane, Carol. “Group Agency and Individualism.” *Erkenntnis*, vol. 79, 2014, pp. 1663–1684, doi:10.1007/s10670-014-9634-9.

San Juan, Valerie, and Janet Wilde Astington. “Bridging the Gap between Implicit and Explicit Understanding: How Language Development Promotes the Processing and Representation of False Belief.” *British Journal of Developmental Psychology*, vol. 30, 2012, pp. 105–122, doi:10.1111/j.2044-835X.2011.02051.x.

Sartorio, Carolina. “Resultant Luck.” *Philosophy and Phenomenological Research*, vol. 84, no. 1, 2012, pp. 63-86.

Schmitt, Frederick. “Joint Action: From Individualism to Supraindividualism.” *Socializing Metaphysics: The Nature of Social Reality*. Edited by Frederick Schmitt. Rowan & Littlefield Publishers, 2003, pp. 129-168.

Searle, John. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press, 1983.

Searle, John. “Collective Intentions and Actions.” *Intentions in Communication*. Edited by Cohen, Morgan, & Pollack. MIT Press, 1990.

Searle, John. *The Construction of Social Reality*. The Free Press, 1995.

Shiffrin, Seana V. “Promising, Intimate Relationships, and Conventionalism.” *Philosophical Review*, vol. 117, no. 4, 2008, pp. 481-524.

Southgate, Victoria, et al. “Infants Attribute Goals Even to Biomechanically Impossible Actions.” *Cognition*, vol. 107, no. 3, 2008, pp. 1059–1069, doi:10.1016/j.cognition.2007.10.002.

Stoutland, Frederick. “Faces of Intention: Selected Essays on Intention and Agency.” *Philosophy and Phenomenological Research*, vol. 65, no. 1, 2002, pp. 238–241.

Surian, Luca, and Alessandra Geraci. “Where Will the Triangle Look for It? Attributing False Beliefs to a Geometric Shape at 17 Months.” *British Journal of Developmental Psychology*, vol. 30, 2012, pp. 30–44, doi:10.1111/j.2044-835X.2011.02046.x.

Tager-Flusberg, Helen, and Kate Sullivan. “A Componential View of Theory of Mind: Evidence from Williams Syndrome.” *Cognition*, vol. 76, 2000, pp. 59–89.

Taumoepeau, Mele, and Ted Ruffman. “Mother and Infant Talk About Mental States Relates to Desire Language and Emotion Understanding.” *Child Development*, vol. 77, no. 2, 2006, pp. 465–481.

Taumoepeau, Mele, and Ted Ruffman. “Stepping Stones to Others’ Minds: Maternal Talk Relates to Child Mental State Language and Emotion Understanding at 15, 24, and 33 Months.” *Child Development*, vol. 79, no. 2, 2008, pp. 284–302.

Thompson, Michael. *Life and Action*. Harvard University Press, 2008.

Tollefsen, Deborah. “Let’s Pretend! Children and Joint Action.” *Philosophy of the Social Sciences*, vol. 35, no. 1, 2005, pp. 75–97, doi:10.1177/0048393104271925.

Tollefsen, Deborah and Shaun Gallagher. “We-Narratives and the Stability and Depth of Shared Agency.” *Philosophy of the Social Sciences*, vol. 47, no. 2, 2017, pp. 95-110.

Tuomela, Raimo. *The Philosophy of Sociality: The Shared Point of View*. Oxford University Press, 2007.

Tuomela, Raimo. *Social Ontology: Collective Intentionality and Group Agents*. Oxford University Press, 2013.

Velleman, J. David. "What Happens When Someone Acts ?" *Mind*, vol. 101, no. 403, 1992, pp. 461–481.

Velleman, J. David. "How To Share An Intention." *Philosophy and Phenomenological Research*, vol. 57, no. 1, 1997a, pp. 29–50.

Velleman, J. David. "Deciding How to Decide." *Ethics and Practical Reason*. Eds. Cullity, Garrett, & Gaut, Oxford University Press, 1997b.

Vetter, Nora C., et al. "Ongoing Development of Social Cognition in Adolescence in Adolescence." *Child Neuropsychology*, vol. 19, no. 6, Routledge, 2013, pp. 615–629, doi:10.1080/09297049.2012.718324.

Walker, Margaret Urban. "Moral Luck and the Virtues of Impure Agency." *Metaphilosophy*, vol. 22, no. 1, 1991, pp. 14-27.

Warneken, Felix, et al. "Cooperative Activies in Young Children and Chimpanzees." *Child Development*, vol. 77, no. 3, 2006, pp. 640–663, doi:10.1111/j.1467-8624.2006.00895.x.

Westlund, Andrea C. "Selflessness and Responsibility for Self: Is Deference Compatible with Autonomy?" *The Philosophical Review*, vol. 112, no. 4, 2003, pp. 483–523.

Westlund, Andrea C.. "Rethinking Relational Autonomy." *Hypatia*, vol. 4, no. 4, 2009, pp. 26–49.

Wilson, George M. *The Intentionality of Human Action*. Standford University Press, 1989.

Wolf, Susan. "Sanity and the Metaphysics of Responsibility." *Responsibility, Character and the Emotions*, edited by F. Schoeman, Cambridge University Press, 1987, pp. 46–62.

Wolff, Robert Paul. *In Defense of Anarchism*. Harper and Row, 1970.

Woodward, a L., and J. a Sommerville. "Twelve-Month-Old Infants Interpret Action in Context." *Psychological Science : A Journal of the American Psychological Society / APS*, vol. 11, no. 1, 2000, pp. 73–77, doi:10.1111/1467-9280.00218.

Zimmerman, Michael J. "Taking Luck Seriously." *The Journal of Philosophy*, vol. 99, no. 11, 2002, pp. 553-576.