THE UNIVERSITY OF CHICAGO

RELYING ON EACH OTHER:
ESSAYS ON TRUST AND MUTUAL UNDERSTANDING

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE ARTS & HUMANITIES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHILOSOPHY

BY

SAMUEL SEGAL

CHICAGO, ILLINOIS

AUGUST 2026

Table of Contents

Acknowledgements

One might expect someone writing a dissertation on reliance to be especially capable of acknowledging all the support they received along the way. I fear I will not live up to this expectation. But here goes.

The biggest debt of gratitude I owe is to the chair of my committee, Anubav Vasudevan. I remember meeting Anubav my first year of graduate school when I took his class on pragmatism and being struck by his calm, kind, and encouraging demeanor. I had no idea then how much of a beneficiary I would be of these qualities over the next nine years. Through all the many dark hours– philosophical, professional, emotional– that inevitably befall a dissertation writer, I could always count on Anubav's wise counsel and steady support. By following his example as a philosopher, I learned to slow down and trust my instincts. I am also especially thankful for all the time he spent with me editing my work; those sessions instilled in me the importance of clarity and precision in philosophical writing.

I am also enormously grateful to Ginger Schultheis. Throughout the dissertation writing process, Ginger was always game to dive into whatever murky idea felt pressing at the time. In all of these meetings, I benefited from her deep understanding of contemporary philosophy, her ability to cut to the core of what I was thinking and subject it to common sense, and her infectious love of thinking. I am especially thankful for the time she spent working with me on what would become the chapter on trust and for her encouragement on that project.

I am also very thankful to have had the input of Anton Ford as I wrote this dissertation. I have always admired Anton's ability to step back from a debate, see the bigger picture, and articulate the broader issues at stake. Every part of this dissertation benefited from that panoramic vision and natural feel for philosophy.

The dissertation developed in big and little ways through conversations with faculty members not on my committee, including Malte Willer, Matt Boyle, James Conant, Agnes Callard, Kevin Davey, and Matthias Haase. It also would not have been possible without the valiant administrative support of William Weaver and Jessica Barbaro.

The feelings of gratitude I find hardest to express are those I feel for my friends and peers. A deep and profound thank you goes to Sandy Diehl, who always manages to be both incisive and inspiring. Another to Michael Powell; the happiest I have been as a philosopher was living with you on Harper. Special thanks, as well, to Ermioni Prokopaki, Melina Garibovic, Andrew Stone, Ryan Simonelli, Maggie Sandholm, Kristen De Man, Kate Petroff, Jacob McDowell, Greg Brown, Harmonie Loritz, John Muller, Stephen Cunniff, and Thomas Willis for being such wonderful people to go through graduate school with.

Finally, I am grateful to my family for their constant support. And to Jiasi for believing in me.

Preface

Human beings are needy creatures. Like all living beings, we rely on the predictable rhythms of the natural world: the sun to rise and light our environment, the weather to bring rain and maintain a steady atmosphere. We rely on various tools to meet everyday demands: clothes to keep us warm, cookware to prepare food, and more sophisticated devices to meet our modern needs. Finally, we rely on each other. We do so in our infancy and youth– in order to mature into capable adults– as well as in sickness and old age; moreover, most of us are committed to lives whose complexity and scope require continuous support from others. These forms of reliance are so deeply ingrained in the human experience that it is hard to imagine it without them. Beings who did not rely– whose success in the world did not depend on things outside their control– would simply be different beings than us.

When dealing with natural phenomena, predictability is essential to our capacity to rely. We may not have direct control over our environment, but its cyclicality allows us to skillfully inhabit it. When it comes to tools, manipulability is the coin of the realm. Tools allow us to extend our agency beyond what is within our immediate control by being sufficiently responsive to our input.

What about persons? Unlike with artifacts and the natural world, there seems to be something misguided about treating persons as entities to be predicted or manipulated. In Strawson's (1962) terms, treating people in this way means abdicating the very stance on them from which they show up as appropriate objects of attitudes like sympathy, gratitude, and care. So while it is possible to treat another human being as a tool or element of the natural world, to do so is to fail to treat them as the kind of thing that they are.

This distinction– between relying on someone as a tool and as a person– parallels one famously drawn by Edward Craig (1990). Craig argues that when someone tells me something, there are two ways I can relate to them as a speaker. On the one hand, I can treat them as a mere source of information, taking their testimony to be a state of affairs whose epistemic significance is ultimately up for me to discern. But I can also treat them as an informant. When I treat a speaker in this way, I recognize them as a subject with a common purpose and trust them for the truth of what they say.

The parallel distinctions that Craig and Strawson draw suggest a more general topic: the distinctive way in which another person can figure in my reasoning about the world. In both the epistemic and the practical domain, reliance on others takes a special form when it involves others. So: what is it to rely on a person? This is the question that the essays in this dissertation take up. The common thesis that runs between them is that reliance on persons requires mutual understanding between relier and reliee. Let me say a bit about what I mean here.

One central concept in contemporary social philosophy is *common knowledge*. Agents have common knowledge of some fact when each of them knows it, knows that they each know it, and so on. Accounts of collective agency, for instance, often stipulate that some fact or set of facts must be common knowledge among the participants. Similarly, theorists of joint attention often claim that the fact that each attender is attending to an object of joint attention must be commonly known. Finally, theorists of communication often assume that some set of communicative principles or maxims is common knowledge among those communicating.

I believe that such stipulations are founded on a genuine insight. There is something about social life *as such* that requires common ground between persons. However, I do not believe that the notion of common knowledge– understood as an unending hierarchy of beliefs–

adequately characterizes this underlying foundation. One of the ambitions of the work that follows is to justify this sweeping claim. Here, we can briefly note one motivation for it, which is that it just does happen all the time that you act and speak with others without consciously engaging in any sophisticated reasoning about what they believe you believe. Accounts of social life which appeal to common knowledge, therefore, seem to risk over-intellectualizing it.

If not through common knowledge, how should we characterize the common ground that underlies our interactions with others? Common knowledge refers to a conjunction of independently specifiable relations, each between an individual and a proposition. In contrast, by "mutual" I mean a relation that two or more agents can have towards an intentional object where each individual's relation toward it cannot be understood independently from the other's. If *predictability* and *manipulability* characterize reliance on nature and tools, respectively, *mutuality*, I suggest, characterizes reliance on persons.

The dissertation is divided into two halves. The first half of the dissertation covers two kinds of grounds practical reliance can have: grounds that are communicative, such as an overt agreement or plan, as well as grounds that are not necessarily communicative, such as a trusting relationship. The second half of the dissertation takes up epistemic reliance. Epistemic reliance, unlike practical reliance, is distinctly communicative in nature. To rely on another person for the truth is to rely on their say-so or testimony. The third and fourth chapters of the dissertation form an extended study of this topic.

In the first chapter, the notion of mutuality plays a key role in my account of trust. The contemporary literature on trust is shaped by the following thought: trust must be understood as reliance plus something else because while trusting someone involves relying, trust but not reliance grounds the possibility of feeling betrayed if let down. My account of trust denies this thought. I argue that trust should be understood simply as the form that reliance takes when properly directed at persons.

I begin defending this view by offering an account of *objectual reliance*. While most writers on trust follow the contemporary literature on reliance in treating it as an attitude one bears towards a proposition, I argue that there is a distinct phenomenon: reliance directed at an object. Just as one can rely on something's being the case, so one can rely on one's phone or one's friend, for instance. I argue that objectual reliance is more fine-grained than and therefore not reducible to reliance on propositions or states of affairs. In particular, objectual reliance is qua-qualified or kind-relative: both the way in which a relied-on object is integrated into one's agency and the kinds of behavior that can be expected from it are constrained by the kind of thing that it is.

Relying on persons, in turn, requires that one rely on a person to behave and to figure in one's action in ways that are characteristic of persons. Trust involves mutuality, then, since it involves the trustee figuring in the action of the trustor in fundamentally the same way as the trustor. Rather than treating the trustee as a force to be reckoned with, the trustor treats the trustee as a coparticipant in the endeavor that depends on the trustee. Thus, I show how the characteristic marks of trust– in particular the fact that trust grounds the possibility of betrayal– can be derived from the very idea of reliance itself. The connection we observe in instances of trust, between depending on another agent performing some action and seeing that performance as owed, is that both of these relations are parts of properly relying on a person. And the intuitive continuity between interpersonal trust and the use of inanimate objects like tools is that both of these are species of reliance.

In the second chapter, mutuality emerges as a structural feature of plan-making. I argue that when agents make a plan together, their plan serves as a common source of each agent's reason to perform their part. I make my case by considering the well-known coordinated attack paradox. Existing treatments suggest that the lesson to draw from the paradox is the importance of common knowledge in bringing about coordination. I argue that these treatments fail because they fail to properly thematize the process by which coordination actually emerges. Even granting that the agents have common knowledge, it is not clear, on these accounts, how coordination would get off the ground.

I identify the root error as a view of plan-making on which the reasons that parties to a plan have to do their parts can be understood independently of their having jointly made a plan. And I challenge this view of plan-making on the grounds that it does not have the resources to explain how, in carrying out a plan, agents rely on one another *as persons.* Finally, I show how we can redeem our initial intuitions about the paradox by rejecting this overly individualistic view and adopting one that highlights the role that social conventions play in making and entering into plans.

In the third chapter, I motivate a certain type of view on testimony, one which treats the interpersonal relationship between speaker and hearer as the source of the reason which the hearer receives to believe the speaker's testimony. I then consider two well-known and persistent problems for this sort of view. The first is a dilemma posed by Lackey (2008): either background considerations of the speaker's reliability do or do not enter into the hearer's mind when they receive testimony. If they do, then such considerations seem to do all the heavy epistemic lifting, and thereby crowd out any supposed interpersonal, testimonial reasons. If they do not, then we seem to get a view of testimony that is too epistemically permissive, one which paints hearers as simply gullible. The second issue for interpersonal views is how to spell out the notion that a hearer's reason for belief is distinctly interpersonal rather than simply another, impersonal or evidential consideration. I conclude that forming responses to these problems is a pressing issue for the interpersonalist.

In the final chapter I chart a path forward by reflecting on the aim of testimony. Most of the literature assumes that the characteristic aim of testimony is generating a belief in the hearer. I argue against this view by considering cases where it seems to matter or be important that a hearer learn about some fact from someone in particular. These cases suggest, instead, that the aim of testimony is something like shared or mutual knowledge. I flesh out this notion of shared knowledge by outlining the role it plays in joint action and by drawing on transparency accounts of self-knowledge.

Finally, I offer an epistemological account of testimony that flows from this view about the aim of testimony, an account I call interpersonal priorism. The transparency of shared knowledge means that parties to shared knowledge have a direct acquaintance with one another's epistemic capacities. They therefore apprehend one another's reliability and credibility as speakers in a distinctly interpersonal way. Drawing on this thought, I suggest that testimonial exchanges are made possible by a relationship of shared knowledge between the hearer and speaker that obtains prior to the testimonial act.

1. Trust as Reliance on Persons

1. Introduction

What is it to trust someone? A natural thought is that trust is reliance: that a trustworthy person is a reliable person and that to trust a person to do something is to rely on them to do it. Though natural, this view is subject to a well-known counterexample.[1] It is said that Kant was so regular in his routine that the residents of Königsberg set their clocks by his daily walks. The residents of Königsberg may have relied on Kant's daily walks, but they did not thereby trust Kant. Indeed, if Kant had broken his routine, the citizens of Königsberg would have had no right to feel betrayed by him. Thus, while trust may entail reliance it seems that it is not "mere" reliance. What, then, must we add to reliance to get trust?

Much of the discussion around trust is structured as a response to this question: it is disputed what extra factor distinguishes trust from mere reliance e.g. an accepted dependence on the trustee's goodwill (Baier 1986), a belief that the trustee has a commitment to doing what they are trusted to do (Hawley 2014), or an expectation that the trustee be favorably moved by the fact that the trustor is relying on them (Jones 2012).[2] In this paper I argue that the assumption on which this dispute rests is false. Trust is reliance. It is reliance on persons.

To see what I mean, consider an analogy with love. Notice that the nature of love depends on the kind of thing that is loved. If I love *that the sun is setting* all that may be required is that I believe that it is good *that the sun is setting*. On the other hand, more is required to love a

[1] To my knowledge, this example was first brought to bear on questions of trust by Baier (1986).

[2] For overviews of the literature which frame it as a debate over what additional feature distinguishes trust from mere reliance see Hawley 2014, Goldberg 2020, and McLeod 2020a. It is important to note that there are an increasing number of writers who have recently argued trust and reliance can come apart entirely. See, for instance, work by Marušić (2017), Thompson (2017), Kelsall (2024), and Faulkner (2025). The concerns of these authors will be addressed in §7 where I consider objections to the view that trust just is reliance on persons.

particular sunset. At a minimum I must be acquainted with *the sunset*. I cannot love a sunset I have never seen. Love of a person is more demanding still. Love of a person arguably requires some sort of interpersonal engagement with the beloved– e.g. the sharing of ends, as suggested by Kyla Ebels-Duggan (2008), or the creation of a collective identity, as suggested by Nozick (1989)– or at the very least the desire for such engagement.[3] Of course, it is possible to love a person merely as something to look at and without pursuing the sort of interaction that is characteristic of love between persons. But to love someone this way would be to fail to love them *as a person*. In this way love is an attitude whose meaning depends on the object that it takes.[4]

I claim the same is true of reliance. Just as love of a proposition is different from love of a person so reliance on a proposition is different from reliance on a person. Once we fully grasp this point, the view that trust simply is reliance– a view that has always been treated as a non-starter– becomes attractive. Trust is reliance on persons *as persons*.

The plan is as follows. In §2 I introduce the most prominent accounts of trust. Most theorists, struck by the connection between trust and reliance but assuming that reliance cannot ground feelings of betrayal, proceed to debate what feature must be added to reliance to account for trust.[5] In §3, I examine the attitude of reliance more closely. Most accounts on offer take it to be an attitude one has towards a proposition.[6] I claim, however, that it is possible for reliance to be directed at objects. One can rely, for example, on one's friend or one's phone. Such *objectual*

[3] Some will disagree of course. For a more permissive view of what counts as love see Setiya (2014).
[4] One could tell a similar story about different kinds of knowing. Kukla (2022) argues that knowing a proposition is different from knowing a city. Knowing a city, in turn, seems to be different from knowing a person since knowing a person, like love, arguably requires some sort of interpersonal engagement with that person. Benton (2025) offers a discussion of knowing a person along such lines.
[5] The literature on trust is obviously vast. In the paper I discuss views of Baier (1986), Holton (1994), Jones (1996), Faulkner (2011, 2017), Cogley (2012), McGeer and Pettit (2017), McLeod (2020b), Hawley (2014, 2017, 2019), Lahno (2020), Simion and Kelp (2023), and others.
[6] For examples of accounts of reliance that treat it as an attitude that takes a proposition as its object see Smith 2010, Alonso 2014, Heape 2020, and Thompson 2017.

reliance, I argue, is not reducible to reliance on propositions. In §4, I use the arguments from §3 as the starting point for a positive theory of objectual reliance. This theory provides the foundation for the view that trust is reliance on persons in §5 and §6. I show how the various features of trust can be derived from the very idea of objectual reliance when applied to persons. In §7 I consider objections. And in §8 I offer concluding remarks.

2. Background: Existing Theories of Trust

Most philosophers agree that trusting someone to do something involves relying on them to do it. But they think trust is normatively valenced in a way that reliance is not. If I trust you to do something I have an expectation that you do it. I feel betrayed if I am let down. Of course, we do sometimes speak of trusting objects, but there is supposed to be an important difference between this broader form of trust and trust among persons.[7] Such "interpersonal trust" is what most philosophers of trust take their topic to be. Contrast interpersonal trust, in this sense, with "mere reliance". Most of the literature takes for granted that to rely on someone to do something is just to act on the assumption that they will do it.[8] Understood in this way, reliance in itself has no normative significance. I can rely on my mail carrier to deliver my mail, for instance, in the same way that I would rely on a machine to do so.

The question of what needs to be added to reliance to account for trust traces back to Annette Baier's (1986) highly influential paper, *Trust and Antitrust*, in which she argues that trust is different from mere reliance in that trust is reliance *on another agent's goodwill*. In a well-known reply, Richard Holton (1994) objects to Baier, pointing out that we can rely on another agent's goodwill and yet not trust them. Consider a confidence trickster who relies on

[7] The common assumption that it is inappropriate to feel betrayed by an object has been questioned, for instance by Nguyen 2022. I return to this issue in §6.

[8] For instance, Hawley (2014) and Jones (2019) explicitly follow Holton (1994) in thinking of reliance as acting on a supposition. See also McGeer and Petit 2017 and Cogley 2012 for similar understandings of mere reliance. We will address this assumption in the following section.

their mark's goodwill. The trickster still does not trust his mark. He cannot feel betrayed if his mark fails to do as expected.[9] Instead, Holton (1994) suggests, 'when you trust someone to do something, you rely on them to do it, and you regard that reliance in a certain way: you have a readiness to feel betrayal should it be disappointed' (4). So Holton posits a certain way of regarding one's reliance on the trustee as what to add to reliance to get trust. Others remedy the issue identified by Holton by supplementing the notion of goodwill. For example, some say that trust also requires a belief that the trustee owes us goodwill (Cogley 2012), while others say that a feeling of shared values is needed in addition to reliance on goodwill (McLeod 2020b).

There are also theorists who think that we need something other than the idea of goodwill entirely to capture the way in which trust differs from reliance. One influential view, developed over a number of pieces by Katherine Hawley (2014, 2017, 2019), is that 'to trust someone to do something is to believe that she has a commitment to doing it and to rely on her to meet that commitment' (Hawley 2014: 10). So, to the notion of reliance, Hawley adds the trustor having a belief that the trustee has a commitment to do what they are trusted to do. More recently, Simion and Kelp (2023) argue that the extra factor Hawley has in mind should be framed in terms of obligations rather than commitments. Lahno (2020), on the other hand, says that shared values are what ground the feelings of betrayal that trust makes possible.

Finally, there are a growing number of theorists who characterize trust as reliance paired with an expectation about how the trustee will respond to the reasons generated by that reliance. Jones (2019), for instance, argues that when I trust a person, 'I count on them to do something and I count on them responding to my counting on them' (957) where "counting on" explicitly

[9] There are other cases where reliance on goodwill does not seem to constitute trust. In a case where I happen to rely on a person's goodwill but perceive their goodwill as paternalistic or out of alignment with my own understanding of my good, it is not clear that I trust that person.

means relying on.[10] Similar so-called "dependence-responsiveness" theories are offered by Faulkner (2011, 2017) and McGeer and Pettit (2017).

There are many important differences between these accounts, but there is one assumption they share: that the way to rule out the kind of case we encountered earlier– Kant's neighbors relying on him for the time– is to posit some feature additional to and understood independently of the concept of reliance. This approach, I suggest, makes their accounts ad hoc. First, they single out interpersonal trust as of special philosophical significance but then, on the basis of one sort of example, reduce it to elements that are not internally related. A more systematic approach would be to seek a better understanding of our original intuition: that there is an important connection between trust and reliance. Such an approach would start with a careful analysis of the idea of reliance.

3. Reliance

I said in the introduction that we can rely on both propositions and objects. But existing accounts of reliance concur that reliance is an attitude one bears towards a proposition.[11] I do not deny that we sometimes rely on propositions. For example, a hiker that sets out confidently without an umbrella relies on the proposition that *the weather will be fine*. This is *propositional reliance*. Is all reliance propositional reliance? In ordinary language we also describe agents as relying on entities: a mountain climber relying on *their rope*, a spouse relying on *their partner*. On their face, these are expressions of *objectual reliance*, expressions in which what the subject relies on is a thing. Of course, no one should deny that we can truly say of someone that they rely

[10] Jones (2019) follows Holton (1994) in thinking of reliance as embedding an assumption into one's planning.
[11] Smith (2010) defines reliance as 'a credal-conative attitude or suite of attitudes that involves both regarding some proposition as true and having a practical commitment or reflectively endorsed attitude towards to[sic] a certain state of affairs' (136). Similarly, Alonso (2014) says that reliance is a practical attitude that serves to cognitively guide our action where 'relying on *p* involves a disposition to, among other things, deliberate on the basis of *p*, plan on the basis of *p*, act on the basis of *p* and draw conclusions from *p*' (166). See, also, Heape 2020 and Thompson 2017 for similar definitions.

on an object. But it is tempting to think such ascriptions are elliptical for ascriptions of propositional reliance.[12] After all, it is not unreasonable to think that a subject, *S*, relies on an object, *O*, to perform some action, ϕ, if and only if *S* relies on the fact *that O* ϕ*'s*. Call anyone who holds this kind of eliminativism about objectual reliance, a propositionalist.[13]

In the remainder of this section I will argue against the propositionalist. In particular, I will argue that the truth of

- S is relying on X to ϕ.

is neither a necessary nor a sufficient condition for the truth of

- S is relying on the fact that X ϕ's.

Consider the following two examples of objectual reliance:

1. I am relying on my map to lead me home.

2. I am relying on myself to be led home by my map.

Suppose, for the sake of contradiction, that objectual reliance reduces to propositional reliance according to the form of reduction given above. Then 1 is true if and only if I am relying on the proposition *that my map leads me home*. Moreover, 2 is true if and only if I am relying on the proposition *that I will be led home by my map*. But the proposition *that my map will lead me* home is identical to the proposition *that I will be led home by my map*.[14] It follows that 1 is true if only if 2 is true. But this claim is false. 1 characterizes my reliance on my map. It suggests that I

[12] My interest here is in whether reliance is an *attitude* an agent can take towards an object. In this sense, objectual reliance should be distinguished from what Smith (2010), for instance, calls "external reliance". My getting to work might depend on the functioning of a part of my car I know nothing about, but this is not a case of objectual reliance, in the sense at issue, since I am unaware of the car part.
[13] The claim that the contents of intentional attitudes can be ordinary objects is admittedly disputed, for instance by Sinhababu (2015). Grzankowski (2016), on the other hand, offers arguments in support of it. That debate aside, my goal here is just to establish as non-trivial the question of how to analyze statements of objectual reliance. I do so by arguing that there is no *simple* way of reducing them to statements of propositional reliance. In this sense, the question of whether every instance of objectual reliance is mediated by an attitude towards a proposition is downstream from the work I undertake in this paper.
[14] I take it that this follows from a straightforward account of propositional identity according to which propositions are identified by sets of possible worlds. Of course, identifying propositions this way is not uncontroversial. For a full discussion of the nature of propositions, which I cannot give here, see Merricks 2015.

have confidence in my map: its legibility, its accuracy, and so on. 2, on the other hand, characterizes my reliance on myself. It suggests I have faith in myself: my ability to read a map, to chart a reasonable course home on the basis of one, and so on.

One might deny that "my map leads me home" and "I am led home by my map" are identical propositions. When the subject is "my map", so this objection goes, the proposition has embedded in it certain information that the proposition with the subject "I" lacks, and vice versa. Perhaps "My map leads me home" implies that *my map* causes me to be led home or that *my map* is able to lead me home. "I am led home by my map", on the other hand, perhaps implies that *I* cause myself to be led home or that *I* am able to lead myself home.

The main insight of our argument, though, can be leveraged against the necessity claim without appeal to propositional identities. When one relies on *a map*, I submit, one takes it to be a good map: a legible, accurate, and, in general, reliable map. Those implications go beyond the information that, let us grant, are embedded in a proposition like, "my map leads me home". To illustrate this point imagine the following case. My enemy gives me a map. It is designed to lead me astray. I am clever, though, and systematically uncover the ways in which my enemy has tried to deceive me: I am able to figure out how to get places by looking at the deceptive map. Thus I confidently set out for a complicated hike relying on the fact that the map will lead me home. So I rely on *the fact that the map will lead me home*. And, let us grant, there is a sense in which I rely on the fact that the map *is able* to lead me home and that the map will *cause* me to be led home. Still, I do not rely on *the map*. It was, after all, designed to deceive me.

I conclude that objectual reliance is not necessary for propositional reliance. We can draw fine-grained distinctions between cases of objectual reliance that we cannot draw with the corresponding cases of propositional reliance.

Objectual reliance is not sufficient for propositional reliance either. Consider:

Two Flashlights

I am camping. It is nighttime, and I need to find my way from my tent to the outhouse. I will need a flashlight so I check to make sure I have one. Luckily enough, I have two: a green one and a red one. When the time comes, I happen to pick the red one out of my pocket, using it to light my way to the outhouse.

In this example, as I head to the outhouse, I rely on the *red flashlight* since that is the one I used to find my way. But I did not rely on *the fact that the red flashlight would light my way*. After all, I had a backup the whole time. Indeed, we could stipulate that I do not even know which one I am using and, therefore, that I do not know that I am using the red one specifically. What is true is that I was relying on *the fact that the red or the green one would light my way*. But it does not sound right to say that I was relying on *the fact that the red one would*. The truth of *S is relying on the fact that X* ϕ*'s*, therefore, is not necessary for the truth of *S is relying on X to* ϕ.

Before proceeding, we need to respond to a regrouping on the part of the propositionalist. Perhaps reducing objectual reliance to propositional reliance is not as simple as we first thought. Still, our responses to the propositionalist seem to arm her with the resources for a more sophisticated version of her own view. One might think that relying on a flashlight, for instance, involves relying on a complicated suite of propositions including the fact that it is a good, reliable flashlight etc.

The first thing to say is that the lesson of Two Flashlights already complicates the task of the sophisticated propositionalist. We saw in Two Flashlights that when it comes to tools, relying on an object means *using* it. This sort of objectual reliance, therefore, requires a complex and continuous engagement with the relied on object. When I rely on a flashlight, my use of it is not so generic as simply relying on *the fact that it is a good flashlight*. I rely on it to light this part of my path, then that part of my path, and so forth. Propositional reliance, on the other hand, is an attitude an agent takes towards conditions that are cleanly outside of their control.[15] Hence, paradigmatic examples of propositional reliance feature an agent on the cusp of action: the hiker setting out for a hike or the commuter aiming to catch a train.[16] The specificity of objectual reliance, then, makes the sophisticated propositionalist's task complex enough to shift the burden of proof.

More importantly, my goal in this section has simply been to get *objectual reliance* into view as a self-standing topic. I have done so by showing that there is no straightforward reduction of it to propositional reliance. It is time for a positive account.

4. Objectual Reliance

Here is a first pass at objectual reliance. Let 'F' range over kinds of objects, such as flashlights and maps. Let 'O' be a name for an object, such as a particular flashlight or a particular map. Then we have:

The Simple Account

An agent, A, (in ϕ'ing) relies on an F, O, (to Ψ) if and only if: O's Ψ'ing is part of how A is ϕ'ing[17]

[15] Smith (2010), for instance, says that an agent can identify the propositional content of their reliance by asking: "given my plan, given my beliefs about the world, given my intentions, and given all my other commitments, what are the remaining conditions that are out of my control but are nonetheless necessary for the successful realization of this plan" (146).

[16] The latter is one of Smith's (2010) examples.

[17] I use the term "part" here in its colloquial sense. I do not mean to commit myself to any mereological theory.

Let us see how the simple account handles some basic cases. Imagine you are driving your car to get to work. Your car running smoothly is part of how you are getting to work. The simple account gets this case right. It says you are relying on your car to get to work and you are. Here is a second case. You are waiting at the airport for a friend to pick you up. She will take you to Penn Station, and you will take the subway home from there. Your friend's driving you to Penn Station is part of how you are getting home. Once again the Simple Account gets things right: you are indeed relying on your friend for a ride from the airport.

Still, the Simple Account is not right as it stands. To see why, consider the following case. Athena is teaching her friend, Nargis, a lesson about the importance of checking the climbing rope when rock climbing. To do this, she allows Nargis to start climbing with a rope that she believes will break once Nargis is a few feet off the ground. Is Athena relying on *the rope*? The Simple Account says that she is: the rope's breaking is part of how she teaches Nargis a lesson. But this sounds odd to say.[18]

Here is a similar example. John is at work in his garden. He notices a weed and goes to pull it out of the soil. But the weed is more stubbornly rooted in the ground than he expects, and he injures his hand in the process. The Simple Account says that John is relying on *the weed* since part of how John is removing the weed is its coming out of the ground easily. Again this is an odd verdict.

Why is it odd to say that John is relying on *the weed* and that Athena is relying on *the rope*? A natural thought is that it is because the weed and the rope are being expected to behave in ways that are not consistent with their characteristic functions.[19] The weed's natural tendency

[18] Arguably, it is less odd to say that Athena is relying on the rope to break. The issue here is that statements of the form A relies on F to Ψ are ambiguous between A relies on *F* (to Ψ) and A relies on *the fact that F Ψ's*. My interest is in the former: situations where an agent relies on an object in virtue of relying on it to do something. For this reason, I treat the question, "does A rely on F?" as a test for whether A relies on F (to Ψ).

[19] I take it that lots of things can have "characteristic functions": my digestive system, a table, food, etc. For a useful recent discussion of the notion of teleology see Graham 2023. I take it to be roughly intuitive that breaking is not

is to remain rooted in the ground. One of the rope's normal functions is to hold weight. If Athena were belaying Nargis, using the rope to hold her weight, then it would sound fine to say that Athena is relying on *the rope*. Similarly, if John tries to sabotage his neighbor's garden by having a weed spread through it, then John is straightforwardly relying on *the weed*.[20]

The fact that whether or not an object is performing its characteristic function affects the aptness of ascriptions of objectual reliance suggests that when one relies on an object, one relies on it *as a certain type of thing*. In other words, objectual reliance is *qua-qualified*.[21] Of course, the claims of objectual reliance that we have discussed so far have not been *explicitly* qua-qualified. But, I claim, they are implicitly qua-qualified. When there is no explicit qua-qualification, a claim of the form 'S relies on an F to V' is interpreted, by default, as meaning that S relies on the F, *as an F*, to V.[22] If we say that Athena is belaying Nargis– and relying on the rope– it is implicit that she is relying on the rope *as a rope*. This accounts for the oddness of saying, without qualification, that Athena is relying on *the rope* when she's relying on the rope to break.

Furthermore, notice that we can make the ascriptions of object reliance that originally sounded odd felicitous by qualifying them. In the case where Athena relies on the rope to break, it sounds fine to say that Athena is relying on the rope *as a torn, frayed object*. And in the case where John relies on the weed to come out of the ground it would be appropriate to describe him as relying on the weed *as something loosely connected to the ground*.

---

one of a rope's characteristic functions. I do not intend what follows to depend on any particular interpretation of this fact.

[20] Of course, there is a sense in which I can rely on a tool to do something other than its natural function. I might use a chair to prop open a door, for instance. I explain how I account for such instances in the paragraphs that follow.

[21] I am borrowing this term from Loets (2021).

[22] One might doubt that all claims of objectual reliance are implicitly or explicitly qua qualified. The cases discussed so far, on their own, do not demonstrate this point. A full defense goes beyond the scope of this paper. Still, the vast majority of examples of objectual reliance are of the form described above: the object relied on is referred to as a certain kind of thing. In such cases, our point generalizes.

Here, then, is our second pass at objectual reliance:

The Refined Account

An agent, A, (in ϕ'ing) relies on an F, O, (to Ψ) if and only if

1. O's Ψ'ing is part of how A is ϕ'ing
2. Ψ'ing is a characteristic function of an F

This definition is an improvement over the simple account for at least two reasons. First, as we have seen, we can explain why Athena, in our earlier example, is not relying on the *rope* when she expects it to fail. Second, we can capture the fine-grained differences that we introduced in § 3 when we distinguished objectual reliance from propositional reliance. Recall, we said there was a difference between relying on the map to lead my home and relying on myself to be led home by the map even though there is no corresponding propositional difference. We can now explain why this is so. According to the Refined Account, these two claims of objectual reliance have different truth conditions. To rely on *the map* to lead me home it must be the case that leading me home is part of the map's characteristic function. On the other hand, to rely on *myself* to be led home by my map it must be the case that it is part of my acting normally to be led home by my map.

Still, the refined account is not right as it stands. As we will see, statements of objectual reliance are doubly qua-qualified. To see what I mean, consider the following case. I am hiking with a friend. We are parting ways but plan to meet up later when it will be dark. I know he will need a flashlight to find his way to me later so I lend him one. Later, he uses the flashlight to find his way to me.

Here, it is part of my way of meeting my friend that the flashlight lights my friend's way. Also, the flashlight's lighting the way is one of its characteristic functions. The refined account therefore says that I am relying on *the flashlight*. But this sounds odd. *My friend* is the one relying on the flashlight, not I.

This case suggests the second way in which claims of objectual reliance are qua-qualified. Relying on a tool like a flashlight requires more than just the fact that its normal functioning is part of how one is doing what one is doing. The flashlight, to be relied on *as a flashlight*, must figure in one's plan or action *as a flashlight*. More generally, in order for a tool to be relied on *as a tool* it must be *used* by the relier. As with the condition that a relied on object must be functioning normally, this second condition may be rather trivial when one relies on something *as an object with a certain property*. If one is relying on a boulder simply *as a heavy object* there may be little restriction on how the boulder figures in one's action.

Here is where we are. Objects are relied on *as a certain kind of thing*. The kind of thing an object is relied on as constrains not only the behavior that is expected of that object but also the way in which it figures in the relier's plan. We arrive at the following account of objectual reliance:

Objectual Reliance

An agent, A, (in ϕ'ing) relies on an F, O, (to Ψ) if and only if:

1. O's Ψ'ing is part of how A ϕ's
2. Ψ'ing is a characteristic function of an F
3. O figures as an F in how A's ϕ's

5. Reliance on Persons

Conditions two and three of Objectual Reliance tell us what it means to rely on an F *as an F*. It follows that to explain what it means to rely on a person *as a person* we need to say something about what it means to function as a person (condition two) and what it means for a person to figure in another's action as a person (condition three).

Let us begin with condition two. What does it mean to function[23] as a person? To answer this question, I suggest we draw on the tradition of thinking about human nature according to which the function or essential activity of a person is to act rationally.[24] What differentiates persons from non-persons, on this way of thinking, is the fact that persons are sensitive to reasons.[25] Why adopt this way of thinking? Consider what makes you the person that you are. Plausibly, it is the normative considerations that guide your behavior: your moral convictions, the roles you take on, the things you value. This dimension of your life is not just one more mark of your human existence– alongside the fact that you take up physical space and digest food. Rather, it is the quintessentially *human* or *personal* dimension of your existence, and it is plausible to think that the capacity to act for reasons is its foundation.[26]

We should also say a brief word on the phrase "acting for a reason". Dancy (2000) draws an influential distinction between normative reasons– considerations which count in favour of an action– and motivating reasons– roughly: reasons that can be cited in the psychological explanation of an agent's action. One simple view, then, is that one acts for a reason when one has a normative reason to so act and one's motivating reason coincides with that normative

[23] When discussing human beings, the phrase "essential activity" may be more apt than the term "function", which some may hear as synonymous with "purpose". I take it, though, that the essential or characteristic activity of an entity, if it has one, can be described in more general terms as its characteristic function.
[24] The roots of this tradition lie in Aristotle's views on the function or *ergon* of human beings. See Korsgaard 2008 for a defense of the relevant parts of Aristotle. See also Matthew Boyle 2012 for a helpful discussion. Obviously, there is room for debate about what the function of a person is. Indeed, we will see that one virtue of the view that trust is reliance on persons is that it helps us diagnose disagreements about trust as stemming from disagreements about personhood.
[25] I am assuming a rough equivalence between the concept of "human being" and of "person". For a view of "personhood" in the same spirit as the one offered here, see Frankfurt 1971.
[26] See chapter 3 of Korsgaard 2018 for an argument to this effect.

reason.[27] My arguments, though, will not depend on any particular account of acting for a reason. I assume that we have enough of a pre-theoretical grasp on this notion to put it to work in our theory.

For a person *to function as a person* is for them to act in accordance with their reasons. To see the appeal of this way of understanding condition two, consider the following pair of cases. Your knee is injured making it hard to walk. You would like to take a painkiller but cannot get up to get one from the bathroom. Luckily, you know that your little brother will do the opposite of whatever you say. Knowing this, you ask him *not* to get you a painkiller from the bathroom. As expected, he brings you the pills. Here, I claim, you do not rely on your brother *as a person* because you do not rely on him to act for the reasons that he has. Rather, you rely on him as an object with certain exploitable tendencies, which are not essential to him as a person. In contrast, imagine you are injured, cannot walk to the fridge, and simply ask your brother to get the painkillers for you. If your brother recognizes that he has a reason to do so and gets you the painkillers, then you are relying on your brother *as a person.*

The following pair of cases illustrates the same contrast. Julius and Katie are walking on a hot day. Julius starts to feel faint and grabs Katie for support. Katie, realizing Julius needs help, positions herself to bear Julius's weight. Julius relies on Katie to act on the reasons that she has– Julius needs support. Here, Julius relies on Katie *as a person.* Contrast this case with the following one. Noula and Rob are watching a movie at home on the sofa. Rob falls asleep and Noula decides to rest her drink on Rob. Here, there is a sense in which Noula relies on Rob. But

[27]According to Arpaly and Schroeder (2014), on the other hand, one acts for a reason when the event of one's action is caused by "other attitudes in virtue of the fact that these attitudes rationalize" (62) the action. See also Williams 1979 and Scanlon 2014 for relevant discussions.

it would be odd for Noula to assert this without qualification. Noula relies on Rob *qua stationary object* but not *as a person.*[28]

These two sets of cases illustrate one key difference between relying on a person *as a person* and relying on a person as something other than a person. To rely on a person *as a person* means relying on them to act on their reasons. To rely on someone to do something they do not have good reason to do or to fail to act for good reason is ipso facto to rely on them as something other than a person.

This completes our account of the second condition of objectual reliance when applied to persons. Turn now to the third condition. When you rely on an object, *as an F*, the object must figure in your action *as an F*. Thus, when you rely on a person *as a person*, that person must figure in your action *as a person*. To see what this means, consider the following case. You are throwing your friend, Arjun, a surprise birthday party. You plan to meet him with all his friends when he gets off work. And you ask his friend Tania to bring the cake. In this case, Tania's bringing a cake and Arjun's being at his place of work are both important parts of how you throw Arjun a surprise party. Moreover, both Tania and Arjun have reason to do what you are counting on them to do; they act *as persons*. Still, I submit, there is a palpable difference between the way in which Tania figures in your throwing the party and the way Arjun in which does. Tania is somehow *more* a part of what you are doing then Arjun is. This is what the third condition of reliance on persons must explain.

Here is what I propose. The reason Tania is *more* a part of what you are doing is that Tania wants to *help* you. She wants your action to succeed. In this sense, she stands in a relation to your action that is similar to the one that you stand in: both of you are active participants in

[28] One might think that there is something morally wrong with relying on a person as something other than *a perso*n. I am also drawn to this thought, but I take no stand on it here.

bringing about the party.[29] Moreover, this way of figuring in your action– namely, as a participant– is not just one way among many for a person to figure in your action. The idea that persons are essentially rational agents suggests that the essential way for a person to figure in an action is as a participant. Arjun, on the other hand, is unaware of what you are planning—indeed, you are trying to surprise him. He contributes merely by being a predictable aspect of your environment. His relation to your action, then, is radically different from Tania's. Tania is a co-contributor or collaborator, while Arjun figures as an object whose spatio-temporal location is to be anticipated.

The intuitive difference between Tania and Arjun provides the basis for our account of the third condition of reliance on persons. When I rely on a person *as a person* I rely on them *to help me* where "to help someone" just means to do something for their or their action's sake. In the example above, Tania helps *you* because her bringing the cake is done to fulfill an obligation *to you*. As I am using the term, then, helping someone need not be a response to an acute need. Helping actions are simply actions done for others.

At this point, we can notice an interesting parallel with reliance on tools. Recall that *use* is the way in which a tool is a part of my action as a tool: a flashlight, for instance, to be part of my action as a flashlight must be made by *me* to light my way. Similarly, for a person to be part of my action as a person is for *me* to be helped by them. In both cases, therefore, there is a sense in which I am the source of the relied on object's behavior. Just as I am the reason the tool does what it does when I rely on it *as a tool* so I am the reason a person does what they do when I rely on them *as a person*.

Let's take stock. We are giving an account of what it means for you, in ϕ'ing, to rely on a person to Ψ. According to my account, three conditions must hold. First, the person's Ψ'ing must

[29] Relevant here is the literature on cooperative behavior. See, for example, Ludwig 2020.

be part of how you ϕ. Second, Ψ'ing must be a characteristic function of *a person* by which we mean the person must have reason to Ψ. Without this second condition we would have no way of distinguishing the two ways in which your brother can be relied on to get you a painkiller. Third, and finally, the person must be part of how you ϕ *as a person*. In particular, we saw that for a person to be part of your action *as a person* is for them to be helping you. Without this third condition we cannot explain the difference between Arjun's and Tania's roles in your plan to throw a surprise birthday party for Arjun.

We can now give our considered account of reliance on persons:

Reliance on Persons

In ϕ'ing, an agent, A, relies on a person, P, to Ψ if and only if:

1. P's Ψ'ing is part of how A ϕ's
2. P has good reason to Ψ
3. One of P's reasons to Ψ is A's good

6. Trust as Reliance on Persons

Now that we know what it is to rely on a person we can turn to our main topic. I claim that to trust a person is to rely on them, in the way required by Reliance on Persons. Call this the Reliance View. In the remainder of the paper I demonstrate the strength of the Reliance View by applying it to the main debates in the literature on trust.

6.1 Betrayal

Let us begin with the example that initially seemed to make the Reliance View a non-starter. Recall Kant whose walks were so regular that the citizens of Königsberg set their clocks by them. Philosophers working on trust have assumed that Kant's neighbors relied on him but did not trust him. Trust grounds feelings of betrayal; reliance, they say, need not.

But this case is not a counterexample to the Reliance View as I have articulated it. The citizens of Königsberg rely on Kant *as an object with certain behavioral regularities*. Perhaps they rely on Kant *as a clock*. But because Kant has neither reason nor the intention to help his fellow citizens by taking his regular walks, the citizens of Königsberg do not rely on *Kant*. They do not rely on him *as a person*.[30]

We can bolster this line of thought by noting that it would be odd for the citizens of Königsberg to approach Kant and tell him: "hey, we're really relying on you". Expressions of reliance directed at persons are normatively loaded in much the same way that expressions of trust are. This point is further evidence for the Reliance View.

The Reliance View is also in a position to account for the fact that trust grounds feelings of betrayal. It says that when you rely on a person, you rely on them to be sensitive to reasons in which you figure. Imagine you are relying on your brother to pick you up from the airport. If you are relying on him *as a person* he must have a reason in which you figure to pick you up. Perhaps he promised to do so. If, then, he fails to pick you up you will be in a position to feel betrayed. You will be in a position to feel betrayed by his neglect of the reason generated by his promise. More generally, when you rely on a person *as a person* you will have legitimate grounds for feeling betrayed if let down: you will have grounds to feel betrayed by their neglect of the reason in which you figure.

6.2 Deception

The Reliance View also accounts for Holton's (1994) example of the confidence trickster, originally posed as a counterexample to Baier's (1986) view that trust is reliance on goodwill.

[30] This case is analogous to another one Baier (1986) considers: shoppers relying on a shopkeeper's concern for profits to motivate him to prevent malicious agents from poisoning his products. My response is the same: the shoppers do not fully rely on the shopkeeper *as a person* insofar as they do not expect the shopkeeper to be motivated, as he should, by the well-being of the shoppers. In fact, Baier's own way of describing the case provides one way of vindicating my response– she says that the shoppers rely on *the shopkeeper's desire for profits* rather than the shopkeeper himself.

The trickster relies on the mark's goodwill in extracting their money. But the trickster does not trust the mark. The Reliance View says, in brief, that for me to trust you is to rely on your sensitivity to reasons in which I figure. This is plainly not the case for the trickster. The trickster knows the mark to be deceived and therefore knows the mark to not have good reason to hand over the money. The Reliance View delivers the right verdict: the trickster does not trust the mark since the trickster does not rely on the mark to be sensitive to reasons that the mark has.

6.3 Trusting at Will

The Reliance View also sheds light on the debate over whether one can trust at will.[31] According to it, whether you trust a person is partly within your control and is partly out of your control. On the one hand, when you rely on someone, you make them a part of how you accomplish your ends. In this sense relying involves making a plan with certain features, which is voluntary. On the other hand, you do not have complete control over the features your plans have. In order to rely on a person *as a person* they must have a reason in which you figure and they must be capable of being sensitive to it. These factors are not entirely in your control. The Reliance View, therefore, explains the complexity of the question of trusting at will.

6.4 Trust as a Two or Three Place Relation

There is also the recurring debate over whether trust is a two place or three place relation.[32] I have taken the three place relation as basic, which is standard.[33] However, the Reliance View has the resources needed to characterize the two place relation. To trust another

[31] Holton (1994) and McGeer and Pettit (2017), for instance, answer in the affirmative while Hawley (2014) and Hieronymi (2008), for instance, answer in the negative.
[32] Many, like Baier (1986) and Hieronymi (2008), say an account should start from the three place relation while there has been dissent from Domenicucci and Holton (2017) and Faulkner (2015).
[33] For this reason I do not directly defend the implication of the Reliance View that trust always occurs in the context of the trustee performing a particular action. Indeed, this is an implication of every view which makes reliance an aspect of trust, which is common.

person full stop is just to treat them as reliable. A similar point holds for explaining the possibility of distrust.[34] On the Reliance View, to distrust is to treat someone as unreliable.

6.5 The Value of Trust

Finally, there is the question of the value of trust. Trust has obvious instrumental value– it helps agents acquire knowledge and undertake complex projects– but articulating its intrinsic value has proved more difficult.[35] One thought is that in trusting someone I express my faith in them and thereby demonstrate respect for them.[36] The Reliance View deepens this explanation. To trust someone is to treat them *as a person.* Since persons *are* persons, trusting someone is a form of recognition.[37]

6.6. Trust, Reliance, and Persons

The idea that trust and personhood are connected is not new.[38] As we saw earlier, Holton (1994) argues that to trust someone to do something is to rely on them to do it and to take the "participant-stance" towards them.[39] In so far as taking the participant stance towards someone means treating them as a person, I am in broad agreement with Holton as well as those who are influenced by him.[40] The contribution of the Reliance View, therefore, is not a completely new account of trust. Rather its contribution is a better understanding of the relationship between trust and reliance. Most theorists think that trust and "mere reliance" are distinguished by an extra

[34] See D'Cruz (2020) for a useful discussion of the issue of distrust.
[35] For a recent discussion of the value of trust see Patrizio 2023.
[36] See for instance D'Cruz (2020) for the argument that distrust can be a way of marginalizing someone and Morin-Martel 2026 for the idea that trusting another involves having respect for their moral character.
[37] Conversely, Brennan (2021) argues that there is a specific kind of trust, "recognition trust", which involves trusting another to recognize us as a person.
[38] Again, See Brennan 2021.
[39] Strawson defines the participant stance as the standpoint "of involvement or participation with others in inter-personal human relationship". See Strawson, 1962, p. 79.
[40] For instance Darwall (2017) draws on Holton (1994) to argue that trust is a second-personal attitude. And Lahno (2020), also drawing on Holton, says that "the first essential element of genuine trust is that we treat the trustee as a person" (p.152). Finally, Berislav Marušić (2017) argues that trust is belief from the participant stance.

factor that is understood independently of the idea of reliance. On the Reliance View, the different elements of trust are internally related to reliance on persons.

Framing the relationship between reliance and trust in this way illuminates how the different parts of trust are related to one another. The conceptual connection between depending on a person's behavior and seeing that behavior as owed is that both of these ways of relating to them are part of relying on them as *a person*. It also bridges a growing divide in the literature by accounting for the relationship between trust and tool use. Nguyen (2022) has persuasively made the point that there is significant continuity between the attitude a rock climber, for instance, takes towards their rope and their climbing partner. In Nguyen's account of this point he goes on to say that one can feel betrayed by one's rope, which is counterintuitive.[41] The Reliance View has a place for Nguyen's original insight while maintaining the integrity of the topic at hand. On the Reliance View, the rock climber relies on both their rope and their climbing partner. Interpersonal trust is simply the form this reliance takes when properly directed at persons.

7. Responding to Objections

One worry about the Reliance View is that it does not account for the fact that trust gives rise to trusting beliefs. At least according to doxastic theories of trust, if A trusts B to ϕ then A believes B will ϕ.[42] But, one might worry, reliance on a person is unlike trust in this respect.[43] Imagine I make a plan to be picked up from the airport by Hans but learn later that he is quite

[41] As a consequence, he rejects the idea that there is an essential connection between trust between persons and the possibility of a morally freighted feeling of betrayal. Unfortunately, I do not have the space for a fuller discussion of how Nguyen's (2022) view compares with mine.

[42] For example of doxastic accounts see, for example, McMyler 2011 and Hieronymi 2008.

[43] A related worry is raised by Berislav Marušić (2017) who argues that trust and reliance come apart because only practical considerations bear on relying, while, *pace* Hieronymi (2008), only theoretical ones bear on whether to trust a person. I deny that relying is simply a practical matter. Relying on *a person* involves conceiving of them *as a person* and thus is partly a function of how one sees them. Moreover, I follow the literature in assuming that trusting someone involves making a plan that depends on them.

flakey. Here, I am not in a position to trust Hans but, it may be suggested, I have no choice but to rely on him.

For one thing, there are many theorists that deny that trust entails belief e.g. in cases of therapeutic trust.[44] If that is right then I *am* in a position to trust Hans: I am in a position where I have no choice but to trust him. The fact that a recurring question about trust– whether trusting a person to ɸ requires believing they will ɸ– can be raised about reliance as well lends credence to our approach. That said, the Reliance View is in a position to meet the objection more directly. I believe what is natural to say, about the case where you have no choice but to rely, is that one has no choice but to rely on the fact that the person will pick me up. If I am really relying on *the person*, this way of speaking suggests that I have faith in the person, treat them as reliable, and therefore believe they will pick me up.

Another important concern is that the Reliance View does nothing to secure any mutual understanding between trustor and trustee. It follows from my theory that I can rely on someone without their knowing it. It might be suggested, to the contrary, that trust essentially involves reliance being mutually recognized or being out in the open between trustor and trustee. It is worth noting that the Reliance View could be adapted to account for this thought. The thought would be that properly incorporating *a person* into one's plans involves that plan being mutually recognized. The concept of being part of someone's plan *as a person* is open to multiple interpretations in a way that mirrors the many choice points a theorist of trust faces.[45] That said, I will briefly say why my intuitions point in a different direction. Consider the act of taking a nap in a crowded park. Here, I take it, one may be "trusting" one's fellow park-goers not to steal

[44] Consider for example parents leaving their house in the hands of their teenager for the weekend who they know may well abuse their trust. Mcgeer (2008) discusses this sort of case as does Carter (2024).

[45] Note here that the Reliance View could also be modified to accommodate the intuition that trust essentially involves commitments. The idea would be that a person is only part of my action as a person insofar as they have a commitment to me to do their part. I am influenced by Darwall (2017) in thinking that trust does not necessarily involve commitments.

one's belongings. And yet, without knowing these fellow citizens, it is hard to see how the fact that this sort of behavior is expected could be mutually recognized. Cases like these, then, diffuse the worry at hand since they suggest that one can trust someone without that person knowing that they are trusted.[46]

A further objection to the view that trust is a species of reliance comes from the thought, which has gained traction recently (Thompson 2017, Marušić 2017, Faulkner, 2025), that trust– the three place relation, A trusts B to ϕ– should be thought of merely as an attitude rather than an activity. Thompson (2017), for instance, considers a case where it may seem an agent trusts another but does not actively rely on them.[47] Imagine you are throwing a dinner party for which a friend, Jane, promises to bring wine. If you happen to have a backup bottle of wine then you may not be actively relying on your friend even though, Thompson argues, you may still trust them to bring wine.

I am not convinced that there is trust in the kind of case Thompson calls our attention to. In the above scenario, if your friend asks you "are you trusting Jane to bring the wine?", it seems like a valid response is: "I *would* trust her to bring wine but there's no need; we have some". Indeed, if you call Jane and say, "I am trusting you to bring the wine" you are implying to her that you are actively relying on her.[48] So there is at least a way of using the word "trust"– A is Trusting B to ϕ– which entails reliance. The lack of ambiguity in expressions of this form, to me,

[46] I do, however, believe that trust, the two place relation– A trusts B simpliciter– does involve mutual recognition. This is a topic for another day.

[47] See Faulkner 2025, chapter 2 for a similar case.

[48] Of course, it is consistent with this point to think that phrases like "A is trusting B to ϕ" ultimately should be analyzed as something like: "A is relying on B to ϕ on the basis of A trusting B to ϕ" where the latter is conceived of as an attitude. An account like that– depending on how one understood trust as an attitude– might not differ greatly from the one offered here. However, it would still raise a question that this paper offers an answer to, namely: what is the conceptual connection between reliance and trust such that we have a particular word for reliance based on trust.

justifies what for a long time was simply an assumption in the trust literature: that without some degree of vulnerability, ascriptions of trust are inapt.

One might press the same objection just considered on more theoretical grounds. Marušić (2017), in particular, argues that trust and reliance cannot be the same thing since they are justified by fundamentally different sorts of reasons. The decision to rely on something, he suggests, is a practical question, while deciding whether to trust someone, he argues, is more like making a judgement; it is not affected by whether it would be expedient to do so.[49]

It seems right to say that deciding whether one can trust someone in general is not affected by practical considerations: they either are or are not worthy of one's trust. But it seems to me that practical considerations do bear on whether one decides to trust someone to do something in particular. We saw this earlier in the case of therapeutic trust. Or consider the fact that it sounds fine to ask a question like, "why are you trusting Jane to get the wine rather than simply getting it yourself?" and that a valid response might be, "because it would be inconvenient to get it myself". So, again, I think there is a strong case to be made against the already controversial view that trust, the three place relation, does not entail reliance.

A further objection to the view that trust is a kind of reliance is offered by Christopher Thompson (2017). He brings our attention to cases where an agent supposedly trusts another but does not actively rely on them. Imagine you are throwing a dinner party for which a friend promises to bring wine. If you have a backup bottle of wine then you are not relying on your friend even though, Thompson argues, you may still trust that they will bring wine. However, I take it that the fact that the natural description is "trusting that" your friend will bring wine rather than "trusting your friend to bring wine" reveals the fact that, without a certain degree of

[49] Marušić (2017) is following Hieronymi (2008) here, although it is worth noting that Hieronymi (2008) does think trust involves reliance.

vulnerability, ascriptions of trust are inapt. Indeed, the idea that trust entails dependence is generally accepted in the literature.[50]

A final case to consider, which one might take to show that there can be trust without reliance, is one where an agent "trusts" another person to do something but for reasons unrelated to the trustor. Imagine I have a friend who has been offered an opportunity to help someone, and let us suppose that she really should take this opportunity and that I make plans based on this assumption. I rely on my friend's sensitivity to reason, but I do not rely on her *as a person*, according to our definition, since their reasons do not feature me. She is not helping *me* in taking the job. Still, one might think, I trust my friend to take the job. My response is two-fold. First, it is not clear that this is a case of trust. As philosophers use the term, trust essentially involves vulnerability to betrayal, and it is not clear that there is such vulnerability here. One might be disappointed that one's friend failed to do the right thing, but feeling disappointed is not the same as feeling betrayed. Second, there may be orthogonal issues clouding our intuitions about this case. It can be a constitutive feature of certain friendships that each be responsive to certain reasons that are otherwise extrinsic to the friendship. I might have a friendship based on classical music, in which case it might be that there is an expectation between myself and this friend that we each do things to support local classical music. In such cases, friends have second order reasons, which directly concern each other, to be sensitive to first order reasons, which do not. If that is right, then in the case under consideration there may well be reliance on persons after all.

One last objection before concluding: why not think that, in the end, the proposed account is still one on which trust is reliance plus something else, namely a certain way in which the relied on object is conceived by the relier. To see how we ought to respond, recall earlier we suggested that certain activities cannot be understood independently of the way in which the

[50] This is one of Baier's (1986) starting points, for example.

object of those activities is conceived of by the subject. We made this point in the introduction when we considered love. Here we can note another example: knowing. Unlike an action like "eating" or "lifting", there is no general act of knowing that can be properly specified independently of how the known object is conceived of by the knower e.g. as a person, a place, or a proposition. On our view, the same is true of reliance. And thus, on our view, the definition of trust is indeed derived from the very idea of reliance when directed at persons.

8. Conclusion

Most of the literature on trust assumes that to rely on someone to do something is just to act on the supposition that they will do it. I have argued that we need to distinguish between propositional reliance and objectual reliance. Objectual reliance– reliance on persons and objects– is qua-qualified. Relied on objects are relied on to perform in a way that is consistent with their nature, and they must be incorporated into the relier's action in a way that reflects the kind of thing they are. In turn, when one relies on a person *as a person*, one relies on the person to act on their reasons and out of a desire to help. This improved understanding of reliance on persons makes possible a powerful and elegant view of trust. We can think of trust as reliance on persons.

2. Coordinating through Plans

1. Introduction

Two generals in the same army have their respective units positioned on opposite sides of a valley. Below them is a city they wish to attack. In order to prevail they must attack simultaneously. An unaided attack by either of them is certain to result in defeat. The generals can only communicate with each other by sending messages across the valley, and each time a message is sent there is a chance that the messenger conveying it will be captured and, as a result, that the message will fail to be delivered.

The scenario just described gives rise to a decision-theoretic puzzle, known as the 'coordinated attack problem',[51] which has been the topic of substantial discussion not only in philosophy but in computer science (Godard and Perdereau, 2021), economics (Binmore and Samuelson, 2001), and game theory (Rubinstein, 1989). On the one hand, it seems like success for the two generals ought to be possible. If they manage to successfully send a series of messages back and forth to one another, it seems that the two generals could thereby come to some agreement as to a time at which they will both attack and, attacking at this agreed upon time, secure victory on the battlefield. To pump the intuition: imagine that the generals succeed in sending thirty messages back and forth, all confirming a time of attack, and that all these messages are written on the same piece of paper so that each time a general receives a new message they can imagine the other general holding the very same piece of paper they are

[51] Initially the problem was conceived of as an illustration of an issue in computer science, namely that of achieving consensus in a multi-agent system (Pease, Shostak & Lamport, 1982) where consensus simply means each agent returning the same value. It turns out that for *n* faulty computers or agents in a given system 3*n*+1 reliable ones are needed to ensure consensus (Pease, Shostak, and Lamport, 1980). In the computer science context the problem is often referred to as the Byzantine General Problem. It is also sometimes referred to as the "Two Generals' Problem" or the "Two Generals' Paradox". An influential presentation of the problem within decision theory was given by Fagin *et. al* (1995) while one of the earliest versions can be found in Akkoyunlu, Ekanadham & Huber 1975.

holding. In such a scenario it is hard to shake the feeling that it would be reasonable for them to attack at the stated time. Yet, obvious as this may seem, a careful analysis of the generals' situation suggests that no attack is possible. The argument, roughly, is this. Since there is always a chance that any given message will fail to be delivered, neither general can ever be sure that a message has had its desired effect of establishing a time of attack. Without knowing, however, that such a time of attack has been agreed to, neither general can be confident enough to risk an attack on their own. They may, of course, seek further confirmation by sending additional messages, but with respect to these messages, the same problem inevitably recurs.[52]

A satisfactory resolution of the coordinated attack problem must somehow resolve these conflicting claims and, more broadly, provide a principled account of the extent to which coordination between the generals is possible. Despite the wide range of theoretical perspectives that have been adopted towards this problem, there are some common assumptions, often left unstated, as to how the problem ought to be framed that apply almost universally across all the proposed resolutions. One such assumption, which will be the focus of this paper, is that the question of whether the generals can coordinate their attack amounts to the question of whether they can each come to occupy an epistemic state with respect to which they are individually justified in attacking at a certain time.[53] In this paper, I argue that this is the wrong way to view

[52] One might think that each confirmation message makes coordination slightly more likely and that, because of this, succeeding in sending messages back and forth could eventually bring about coordination by making each general gradually more confident of an attack. The problem is that it is not clear why each additional message should make the generals more confident of an attack. Each general, if they are rational, will not blindly attack after receiving a certain number of messages but only attacks if they believe the other will. Moreover, the generals commonly know this about one another. So each knows that the other will only attack if the other fully believes they are going to attack as well and thus not simply in virtue of receiving a certain number of messages. Of course, if either general is irrational or believes the other is irrational, then additional messages may have this sort of gradual effect. Lederman (2017) explores such scenarios in a response to the problem I discuss in the next section.

[53] For some recent treatments of the coordinated attack problem– to be discussed in more detail in the next section– see Lederman (2017), Paternotte (2017), Harris (2020), Thomas (2021), and Sbardolini (2024).

the problem and that its successful resolution requires that we reframe the problem as one of collective, rather than individual, agency.[54]

In Section 2, I criticize existing discussions of coordinated attack on two related grounds: first, I argue, they leave opaque the mechanism by which coordinated actions are presumed to occur; and, second, they are not clear about the conditions on adequate solution to the problem. In Section 3, I diagnose these shortcomings as the byproduct of a widely held, but ultimately mistaken view about how it is that agents, through the relevant sort of communicative exchange, become rationally entitled to do their parts in a plan. I call this problematic view of plan-making the *separatist* view. It asserts that the reasons that parties to a plan have to do their parts can be understood independently of their having jointly made a plan. One immediate consequence of the separatist view is that an agent cannot, on pain of circularity, cite their participation in a plan as itself a reason as to why they ought to do their part in that plan. But such intrinsic reasons, I argue, are essential to the very idea of coordination. To highlight this point, I explain how the sort of coordination that we intuitively take to be attainable by the generals is precluded by a separatist view of plan-making even when they are allowed to communicate via a perfectly reliable messaging system. In Section 4, I offer an alternative to the separatist view of plan-making– the *interpersonalist* view– which emphasizes the role that social conventions play in making and entering into plans. Finally, in Section 5, I revisit the coordinated attack problem in light of the earlier discussion. Existing attempts to resolve the problem, laboring under the

[54] In suggesting we solve the problem of coordinated attack by rejecting an overly individualistic picture of decision-making, my approach invites comparison to theories of team reasoning as developed, for example, by Robert Sugden (2000, 2003, 2018) and Michael Bacharach (1999, 2006). The difference between the solution to coordinated attack I propose and these theories will quickly become clear. I address it at greater length in Section 3. Briefly: the view of plan-making I offer and the idea of team reasoning, as developed by Sugden, Bacharach, and others should be seen as two distinct amendments to traditional decision theory. The former concerns how communication between agents can rationalize certain behaviors that would otherwise be irrational while the latter concerns situations where no communication is needed and where agents are simply presented with a fixed set of possible actions.

constraints of the separatist view, have attempted to complicate or qualify the nature of the individual epistemic achievements that underwrite coordination in an effort to allow the generals to agree upon a plan of attack. This, I argue, is a misguided approach. The generals can agree upon a plan of attack simply by satisfying a plan-making convention. The fact that there are no in principle barriers to them doing so points to the resilience of social conventions in conditions that are not ideal for communicating.

2. Discussion of Existing Views of Coordinated Attack

The classical approach to coordinated attack is to reject the intuition that the generals could ever rationally perform a coordinated attack (Halpern and Moses 1990, Fagin et al., 1995, Harris 2020). There are two key assumptions made in carrying it out. First, it is assumed that the generals, because they are rational, would each need to know that the other was going to attack in order to attack as well.[55] Second, it is assumed that the generals have common knowledge that they are rational and thus that they commonly know that each would only attack if the other was going to attack.[56] These assumptions taken together are drawn on to show that the reason the generals cannot coordinate is that they cannot establish common knowledge that they are going to do so. The argument proceeds in two stages.

First, assume, for the sake of argument, that the generals were planning an attack. Then, following the first assumption, each general would need to know that the other general was going

[55] What I want to highlight and eventually subject to scrutiny is this interpretation of what rationality requires of the generals, not that they are rational full stop. It is worth noting, though, that in making this assumption we bracket a valid question to ask at this point: how do actual people respond in situations like the ones the generals are in? Interestingly, empirical studies show many people doing the equivalent of attacking after receiving a sufficient number of messages. See for example, Thomas et al. 2014. While these studies are valuable, in that they force us to think in more nuanced ways about actual human psychology, they do not furnish us with any clear solution because they do not directly engage with the argument that purports to show that it is not rational for the generals to coordinate. They do, however, make a solution that shows how the generals could be rational in attacking more desirable.

[56] A group of agents have common knowledge of a proposition just in case each of them knows that proposition, each of them knows that each of them knows it, each of them knows that each of them knows that each of them knows it, and so on ad infinitum. This definition can be traced back to Lewis (1969) although his analysis is put in terms of belief.

to attack. Additionally, following the second assumption, each would know that the other general would need to know this fact, and each would know that they know this, and so on and so forth. So if the generals were attacking then they would have common knowledge that they were going to attack.

The second part of the classical approach is to argue that the generals could never attain such common knowledge. Notice, the generals do not start out with common knowledge that they are going to attack. Clearly, the only way for them to acquire it is by sending messages back and forth. But imagine what would happen if they tried to establish common knowledge in this fashion. A could send a message to B indicating *that they'll attack at a certain time*– call this fact *p*. The message having gotten through, B will know *p* upon receipt. So A and B can know *p.* But A is not certain their message has been delivered so A does not know that B knows *p*. B, realizing this, could send a message back to A. If they get lucky again and this confirmation message is received, A would know B knew p. After two messages, though, the generals would still be woefully far from common knowledge of *p*. They would know *p* and they would know that they know it. But B, not knowing the latest confirmation message was received, would not know that A knows that B knows *p*. It should be clear by now that any attempt to establish common knowledge by sending a confirmation message is futile: for any *n* number of messages the generals successfully send they will have at most *n-1* layers of reflexive knowledge of *p*.

This concludes the reconstruction of the classical account. A less than fully reliable messaging system prevents the generals from establishing common knowledge, but the generals would need common knowledge in order to attack. Thus, no attack is possible. What we have learned, on this reading, is that our initial intuition– that victory should be possible– was not sufficiently sensitive to the deep connection between common knowledge and coordination.[57]

[57] For work exploring this connection see Greco 2023 (Chapter 7) and Sbardolini 2024.

The influence of the classical account is shown in the fact that the coordinated attack problem is strongly associated with this supposed connection. In many cases it figures simply to motivate issues related to common knowledge, which could be raised independently of it, for instance, the question of whether and how common knowledge can be psychological reality among cognitively finite creatures (Clark and Marshall 1981; Lederman 2018; Thomason 2021). Or, to take another example, there is the question of whether a fact being common knowledge is equivalent to its being public or out in the open among agents (Paternotte 2011).

Recently, there has been pushback against the classical account (Lederman 2017); Paternotte 2017). Notably, Lederman (2017) questions the assumption that the generals commonly know that they are rational, which may be unrealistic. What creates the puzzle, after all, is an intuition about what it would be rational for the generals to do, but this is arguably an intuition about what would happen in the real world, and real world agents rarely commonly know one another to be rational. If we drop this assumption, the argument given earlier, which showed that common knowledge is necessary for a simultaneous attack, no longer works. Lederman, therefore, makes space for a response to the problem that embraces the intuition that coordination be fully rational after a finite number of messages.

In the remainder of this section, I want to raise two problems for the literature on coordinated attack, which apply equally to both sides of this debate. Here is the first. Coordinated attack presents two agents and asks us whether they can rationally undertake a certain set of actions, a simultaneous attack. Answering in the affirmative should involve spelling out what a deliberative process concluding in their attack would look like. Answering in the negative should involve saying why such a deliberative path is blocked in this case. The issue that is never made sufficiently clear is just that. How are we to imagine each general *transitioning from* or *failing to transition from* not having sufficient reason to attack at a certain time to having sufficient reason to do so? Naturally, the generals' only hope is to communicate with one another by sending messages. But this observation is clearly only the beginning of the story. How exactly is it that a communicative exchange could be the intermediary between the generals being undecided about what to do and them undertaking a simultaneous attack?

To the extent this issue is addressed, it is suggested that it does so by allowing the agents to form common knowledge of the fact that they are going to attack (Fagin et al. 1995; Lederman 2017; Harris 2020). But this cannot be right. This fact– the fact that the generals are going to attack– takes for granted that A and B have already decided to attack, and it is just their deciding to attack that we are trying to understand. While common knowledge may be a necessary condition on coordinated behavior *once it is underway*, this observation should not be confused with an account of how it is that agents get coordination off the ground. One could frame this point as a regress problem. If agents are understood as coordinating by acquiring common knowledge of the fact of their future coordinative behavior, this view would only reraise the question of how such a fact originally obtained. The problem with the view under consideration, then, is shown by the awkwardness in specifying the content of the generals' messages to one another. The content of the first message is typically taken to be the proposition that the generals require yet cannot achieve common knowledge of. But it is awkward to imagine, as we did earlier, one general bluntly messaging the other, "we attack at dawn tomorrow" as opposed to the more natural, "let's attack at dawn". Indeed, most theorists who take the standard approach gloss over the content of the generals' messages to one another.

There are analysts of coordinated attack that posit the importance of common knowledge of a fact which does not itself imply that the generals are already planning an attack. That is, it is suggested that what is crucial is common knowledge of some fact which could legitimately precede the generals deciding to attack. Here, the awkwardness is not with spelling out the content of the message but spelling out the content of the common knowledge. For instance, analysts who go down this route have suggested the generals require common knowledge of facts as various as: the fact that an instruction has been given to attack at a certain time; the fact that the enemy will be unprepared at a certain time (Morris and Shin 1997); the fact that a proposal has been made to attack at a certain time (Binmore 2008); the fact that one general is saying: let's attack at a certain time; the fact that a message has been sent (Chwe 2001); or the fact that a message proposing to attack at a certain time has been delivered (Halpern and Moses 1990). The problem here is that common knowledge of any of these facts would not rationally entitle either general to attack at the specified time. This is because none of these facts nor their being commonly known by the generals imply that either general is going to attack at a certain time. General A can reasonably think, "B has instructed me to attack and knows that I know this and so on, but is B really going to attack?". Common knowledge of facts which fail to entail that either general is going to attack would ipso facto fail to sufficiently motivate either general to attack.

The first problem with the way coordinated attack is often discussed is that an explanation of how a communicative exchange could motivate or, by failing to obtain, could fail to motivate the generals to attack is never given. Common knowledge, contrary to what is often assumed, is not up to this explanatory task. The generals can either have common knowledge of a fact which entails they have already decided to attack, in which case the account begs the question, or it will be of a fact which does not entail that the generals will attack, in which case such common knowledge will be motivationally insufficient.[58]

The second problem I want to raise for these discussions is that they are not explicit about what would count as a solution to the puzzle. Both sides of the debate around coordinated attack discussed above assume that the question we are trying to answer is whether it could be rational for the two generals to attack simultaneously. But the problem is supposed to be a coordination problem. So both sides assume that the generals attacking simultaneously is just no different than them performing a *coordinated* attack. But is this right? To see why not, consider the following case.

The generals' system of messaging one another is fully reliable. Moreover, A has a way of sending a message to B that proves beyond any doubt that A is going to attack at a certain time. Perhaps A can send a video to B, ordering their troops to attack at a certain time, which is taken as a point of no return. In light of this, A realizes they can serve their interests perfectly well without making arrangements with B. A can simply decide to attack, inform B of their settled plans and rely on B to do what is in B's interest, namely to follow suit. Suppose A does this and is successful. So instead of establishing a plan with B, A simply acts unilaterally, and B, having no other option, matches their attack.

In the above scenario the generals confront a coordination problem and, ultimately, end up realizing their mutual interests. Moreover they do so rationally. We can assume that B's only rational response to A's message was to attack at the same time as A and that A knew this and was therefore perfectly rational in their approach. Does it follow that the generals performed a coordinated attack? On the one hand, general A foresees the advantageousness of a simultaneous

[58] One might worry that we have proved too much since the argument seems to show that no amount of knowledge can motivate a coordinated attack. I believe this is a consequence to be embraced. On the view I ultimately espouse, entering into a coordinative plan is as much a practical achievement as it is a theoretical one.

attack and successfully contrives to bring it about. Thus, one might say, it is no accident that the generals' behavior ended up producing the result that it did. But behavior flowing from clearly antagonistic relationships, such as the complex strategies of two poker players, is also non-accidentally related. These dynamics, too, produce effects that are foreseen by both parties. At least on a fairly natural reading of the term, however, the generals' behavior in the foregoing example as well as that of the poker players' fails to be coordination.[59] This is the sense at play when we talk about coordination as something agents do or accomplish, the sense that applies to agents that are coordinating *with* one another. In the case above, A acts unilaterally while B must passively conform to A's vision. This dynamic, where one agent forces the hand of the other, is so one-sided that it belies the idea that A and B have coordinated with one another.

The example considered above suggests a distinction between two kinds of states agents confronting a coordination problem can be in: the state in which the each agent is rationally entitled to *coordinate* with the other and the state in which each agent is merely rationally entitled to do their part in a scheme of behavior that is to their mutual benefit. In the next section I provide a principled explanation of this distinction. For now my purposes are served if it is acknowledged that this issue– what exactly counts as coordination– is something that we need to confront in order to properly respond to coordinated attack. The problem, recall, stems from the intuitive sense that the generals are in a position not just to attack simultaneously but to coordinate their attacks by sending messages back and forth. A satisfactory response, whether it embraces or reverses this intuition, must, for this reason, take a stand on what coordination is. Most discussions, however, fail to make contact with this issue.

[59] Relevant here is the distinction between coordination and cooperation, a discussion of which can be found in Ludwig 2020. I do not mean my argument to turn on a particular definition of coordination. The above argument could be reformulated in terms of cooperation or simply in terms of plan-making. What is essential is that the initial scenario generates in us the intuition that something more than mere simultaneity is possible for the generals. We feel not just that they should be able to attack but, more strongly, that they should be able to agree on a time of attack. Because it is known as the problem of *coordinated* attack, I couch my argument in terms of coordination.

3. Diagnosis

I have argued that in discussions of coordinated attack there is a lack of clarity about two related issues: on the one hand, the mechanism by which agents decide to coordinate, and, on the other, what it means for agents to coordinate. In this section I want to identify the root cause of these issues: a questionable theory of how agents, in communicating with one another, become rationally entitled to do their parts in plans.

Imagine you and I decide to meet for coffee at a certain time and place. Doing so presumably requires some sort of contact between us through which we work out the plan's details: the actual time we will meet, the specific place, the expectations we will have for our meeting. Before this contact neither of us had a particularly strong reason to go to the specified time and place. After our meeting we each acquire decisive and determinate reason to do so. How is it that simply uttering some phrases back and forth performs this rational alchemy?[60] This question is central to understanding the coordinated attack problem. The initial intuition that success for the generals is possible is rooted in the sense that they could agree on a time of attack. Indeed, plans or agreements are a paradigmatic way in which agents coordinate themselves.[61] Thus, to adequately respond to the seemingly attractive argument that coordination between the generals is impossible, we need a stronger sense of why we originally thought plan-making was possible. We need to understand how an exchange of words can rationally entitle agents to enter into a plan.

Let me say a few words about how I am conceiving of this topic. As far as I can see, agents having a plan does not necessarily entail that those agents are acting together or engaged

[60] One way of hearing this question is as a question about the illocutionary force of speech acts like proposing and accepting. The literature on illocutionary force tends to focus on acts like assertions, commands, and promises, but acts like proposing and accepting seem to constitute an equally valid category of speech act. For a good overview of this literature see McDonald 2022.
[61] For an insightful study of the connection between plans and shared agency see Jules Salome-Sehr 2024

in a joint project. Corporations make plans when they make a deal with one another, but obviously such corporations are not seeking to embark on a shared project.[62] Thus, agents may enter into such a plan without doing any planning or collective deliberation. Furthermore, I will treat the question of how agents "enter into plans" as tantamount to the question of how each agent acquires sufficient reason to do their contribution to the plan. I will call this contribution an agent's "part." Thus I want to distinguish reasons for parts, which I discuss below, from any marginal or trivial reasons the parties may have prior to making their plan[63] as well as any reasons parties to a plan may acquire subsequent to their making the plan. For instance, parties to a plan may have reason not to let one another down. Or, similar to the way in which one might think individuals have reason to do what they have planned to do over and above their initial motivating reasons,[64] one might think that parties to a plan have additional reason to do their parts in addition to their having made a plan in the first place. My interest, however, is just in exactly those reasons that agents who enter into a plan acquire by making a plan.

A crucial division in the space of possible views is between ones that I will call interpersonalist and ones I will call separatist. What distinguishes these camps is a normative issue, namely the precise source of the reasons for parts parties to plan inevitably acquire. The interpersonalist says that in making a plan with one another, agents thereby, as a consequence of their communicative exchange, give themselves reason to do their parts and require no further plan-independent reasons. Notably, for the interpersonalist, for a given party, there need be no

[62] This distinguishes plans from collective intentions as well as from episodes of team reasoning. For just a few key examples of the extensive literature on collective intention see Searle 1990, Gilbert 1989, and Alonso 2009. A discussion of the connection between collective intentions and team reasoning is offered by Gold and Sugden (2007).

[63] I make this caveat since one might think that each general has a very weak reason to attack at a given time even before any messages are exchanged since there may be a small chance the other will happen to attack this time. The existence of such weak, ubiquitous reasons is defended, for example, by Schroeder (2007, chapter 5).

[64] For a defense of this position see Gauthier 1994, and for useful more recent discussion of this issue see Smith (2016). It is also discussed in Bratman (1987).

assurance that the other parties will do their parts prior to entering into the plan. Rather, each party's rational entitlement depends on and flows from a common, interpersonal source. The separatist says that reasons for parts can and should be described independently of the idea of a plan. The mechanisms by which each party acquires reason to do their part are therefore to be understood separately, as each agent acquiring individual justification for their part. A paradigmatic example of a separatist view, therefore, would be that each party to a plan is motivated to do their part because they come to believe each other party will do theirs.

An analogy with individual action will help clarify the divide. It is interesting to note that individual actions which are composed of distinct sub-actions arguably require a kind of *intra*personal coordination.[65] Consider something as simple as an agent taking a cab to the airport in order to catch a flight. For such an agent, it would be unreasonable to plan or expect to fly if they were not planning to take a cab to get to the airport. And conversely it would be unreasonable for them to take a cab to the airport if they were not planning to fly once one got there.[66]

An agent's performing such a simple sequence of actions is easily comprehensible, but performing them requires confronting a situation much like the one the generals confront. Consider the moment before the agent gets into the cab. Doing so is not rational if they do not know that they are later going to get on the plane. But it is not clear what their grounds for such surety could be. It is not true that in every instance in which an agent pursues a complex plan that they always complete it. So the agent getting in the cab certainly has no inductive grounds for

[65] For a more extensive discussion of this topic than I can give here see Heeney (2020) and Ferrero (2022).
[66] The point here does not have to do with certainty. You might think that such an agent only needs to be 90% confident that they are going to follow through on their plan. Even granting this, the problem remains that they need to be 90% confident about what their future self is going to do, and that, it seems, is not obviously in their present control.

believing that they are going to follow through on their plan. In absence of any such guarantees individual agents still constantly manage to coordinate with their future selves.

We can make sense of this fact if we assume that agents are able to make decisions about what they are going to do now and in the future.[67] That is, we can say that an agent prior to going to the airport can decide all in one moment to perform a complex sequence of actions: getting a cab in order to get on a plane.[68] Of course, there might be situations where agents do things like go to the airport and where that action is sustained by a series of separate episodes of deliberation. Still, common sense suggests an agent can also decide all in one moment to perform an action with a sequence of stages. The interpersonalist thinks there is a similar deliberative possibility available to multiple agents, one that does not coordinate self and future self but self and other.

For many philosophers the separatist picture operates more or less in the background as the default view.[69] Two exceptions– Abraham Roth (2014)[70] and Margaret Gilbert (1989, 1993)– are worth pointing out.[71] Roth's view is that, in cases of shared activity, it is possible for one agent to "take up and directly act on another's intention". Insofar as one agent directly acting on

[67] I intend a relatively modest claim here: that an individual agent can make a single decision about what they will do now and in the future. I take this claim to be consistent with a range of positions on a set of related and contentious debates in the literature on diachronic agency. Notably, there is the question of whether there are irreducibly diachronic norms of agency, a good summary of which is given by Doody (2019). Also relevant is the question of whether facts about what one actually is going to do in the future are relevant to the question of what one should do in the present. An overview of this issue is given by Zimmerman (2017).

[68] Korsgaard (2009) makes this point in order to show that the reasons one acts on in a given instance must be reasons one can act on in the future.

[69] The fact that separatism operates as the default view for many philosophers is exhibited in the assumption we highlighted earlier that is made in many discussions of coordinated attack, namely that the question of whether the generals can attack simultaneously is just the question of whether each can come to believe that the other is going to do so. I take it that it is motivated by a skepticism about the possibility of essentially interpersonal reasons.

[70] Roth's focus is slightly distinct from mine here since Roth is interested in joint actions while plans, as I have argued, do not necessarily have to issue in a joint action. Still coming together to act jointly is an important example of plan making and this means that Roth, in my terms, holds an interpersonalist view.

[71] See also Velleman (1990) and Julius (2013) for related views. I take it that de Kenessey (2020, 2022) can also be read as offering an interpersonalist picture of agreements but this would hinge on how exactly one understands the sense in which joint decisions give agents reasons. For reasons to be discussed in the next section, I do not read team reasoning theorists like Sugden (2000, 2004) or Bacharach (1999, 2006) as offering anything like an interpersonalist view of plan-making.

another's intention generates reasons for those agents to do their parts in a plan, Roth's view is an interpersonalist one. Another interpersonalist picture of plans is articulated in the work of Margaret Gilbert. Gilbert asserts that agreements, the word she uses for what I call plans, are not created by two individuals making what she calls a "joint decision". Again, in so far as such joint decisions generate decisive reasons to do parts, Gilbert also offers an interpersonalist picture of plan making. Having laid out these two camps, my plan for the rest of the section is as follows: I establish that discussions of coordinated attack assume the separatist view, and I argue that the obscurities in their accounts are symptomatic of this commitment.

The fact that the majority of discussions of coordinated attack assume a separatist picture of plan making is borne out by the fact that these discussions– about what is rational for the generals to do– turn exclusively on the question of what the generals are in a position to know. Only if we assume the separatist picture is this a reasonable limitation. As we said, on the separatist picture the significance that a plan making conversation has for the agents is not that it conjures some independent thing, a plan but rather is connected to how it influences what each agent knows. These epistemic inputs then become the material for each agent's own deliberative process. If the interpersonalist is right, on the other hand, questions of what it is rational for the agents depends not only on what they *know* but also on what they *do*, that is whether they in fact create a plan.

A related way in which the separatist picture can be seen to be at work in most discussions of coordinated attack is that they fail to consider the range of effects different messages might have on the generals. All ways of approaching the problem have to make certain assumptions about what happens when a general receives a message. Even if the message is a simple statement of fact– such as, "the enemy will be vulnerable tomorrow"– there's a question as to what its effect will be. Does the receiver simply then believe *the enemy will be vulnerable tomorrow*, do they believe *the other general sent a message saying "the enemy will be vulnerable tomorrow"*, or do they form some more complicated set of beliefs? In the same vein we could ask: what happens when a general receives a message whose content is not factual, such as: "let's attack tomorrow at dawn" or "okay, sounds good"? It is not unreasonable to think such messages may have effects that are not simply epistemic in the sense that they directly alter the normative relations between the speakers. The influence of the separatist picture on this literature can be gleaned from the fact that this kind of issue, despite its relevance, never comes to the fore.

Finally, it is worth pointing out that most analysts of coordinated attack would likely welcome the ascription being made here. Much philosophical interest in coordination problems was inspired by Lewis (1969) who wanted to show that solutions to coordination problems could arise without any explicit agreement between the agents. In turn, much work on coordination problems has been connected to research programs which seek to explain collective intentionality in terms of individual intentionality.[72]

So, what are the downstream effects of the separatist picture on accounts of coordinated attack? I will argue that such a picture prevents us from representing agents, in situations like the one the generals are in, as genuinely coordinating with one another. What creates the problem is a structural feature of the generals' practical situation: it only makes sense for general A to attack at a given time if general B attacks at that time, and it only makes sense for general B to attack at a given time if A is going to attack at that time. To put a label on this issue, we could say that the

[72] See for example: Gauthier 1975; Aumann 1987; Vanderschraaf 1998; Cubitt and Sugden 2003; Sillari 2008. Of course, these authors also have many important, nuanced points of disagreement. See, for instance, Cubitt and Sugden's (2003) critique of Aumann (1987) and Vanderschraaf (1998).

set of actions that are to their mutual benefit exhibit rational interdependence.[73] This is a non-trivial obstacle. It seems like each needs to go first, but the two of them are in perfectly symmetrical positions, each waiting for the other. In some such cases, social protocol breaks the deadlock. Drivers choosing which side of the road to drive on confront rational interdependence, but traffic laws thankfully solve that dilemma. In other cases the fact that a certain set of actions is salient among the agents may be relevant.[74] The generals, on the other hand, have only themselves to fall back on. What resources does the separatist have to offer? It is a hallmark of their position that at any time at which one general acquires reason to do their part we can ask whether the other general already had their reasons in place. It is precisely this sharp division between their rational processes that the separatist maintains. A consequence of this division is that at least one general must decide what to do without knowing what the other general will do. The issue is that doing so would seem to be irrational. On the separatist's behalf, I will consider and ultimately reject two ways around this issue.

One way around the problem of interdependence, for the separatist, is a solution like the one we considered in the previous section. One general acts unilaterally, limiting the set of rational responses available to the other so that they are, in a sense, forced to attack at the same time. In this type of case, one general gains the ability to act rationally by exercising control over the other general and the other general can be rational in how they cope. To motivate moving away from the separatist picture of plans, then, we need to say more about why this fails to be coordination in any meaningful sense. At first glance it is the presence of some sort of asymmetry between A and B. One agent is literally imposing a certain scheme of behavior on the other and the other is forced to passively conform to it. What sort of asymmetry is at stake here?

[73] This is a well recognized pattern in coordination problems such as Hi-Lo and Stag Hunt. It is referred to by Colombo and Guala (2023) as a circularity problem.
[74] This is a well-known phenomenon theorized by Schelling (1960). For a more recent overview see Schönherr 2021.

A and B perform identical actions. So it cannot be a matter of the acts that make up their supposedly coordinated behavior. Rather what we are picking up on here, I propose, is an asymmetry between how each agent relates to the other agent's way of making a decision. In order for A's strategy to work A has to conceive of B, not as free to choose whatever course of action they like, but as someone to be manipulated and whose action can be predicted in advance of their having freely settled what to do. A's strategy, that is, only works if A is sure when A chooses to attack at a certain time that B will follow suit. B on the other hand cannot relate to their own action in this way.[75] For B it is an open question whether to attack up until the time when B has decided to attack. Notice there is the inverse of this asymmetry between how B thinks of A's action and how A does. For B, A's action shows up as an inevitable feature of the world that must be dealt with. A on the other hand, relates to A's action as we said B relates to B's: as something to be freely chosen. What makes this case not one of genuine coordination are these two mismatches between, on the one hand, A's perspective on A's action and B's perspective on A's and, on the other, B's perspective on B's action and B's perspective on A's. Avoiding such a mismatch is a relatively straightforward condition on coordinative behavior, I will call this the <u>symmetry constraint</u>. So far the separatist still lacks an account of how the generals could overcome the problem of interdependence that meets the symmetry constraint.

One might think one could give such an account by drawing on the idea of a conditional promise.[76] Perhaps one general could promise to attack at a certain time on the condition that the other promises to do the same; then, the other could express a promise to do so, causing the first promise to go into effect. There are two reasons that this proposal falls short. First, plans and

---

[75] An exploration of the distinction between seeing your future actions as predictable and seeing them as open is given by Marušić (2015, chapter 6).

[76] The problem I pose for the separatist view of plan-making is similar to what is known as the "Settle Problem" in the literature on collective intention, originally raised by Velleman (1997). Rachar (2023) offers an approach to that problem which seeks to resolve it by appealing to conditional intentions.

promises are too closely related for the purposes at hand. As de Kenessey (2020) points out, when a promise is accepted, the promisee, as much as the promisor, acquires reasons for action–minimally, to not interfere with the promisor's fulfillment of the promise. Pointing to promises as an explanation of plans only postpones the task of saying how it is that agents, by communicating, can generate simultaneous practical reasons.[77] Second, more importantly, plans have a further dimension of symmetry that this proposal fails to capture. When agents have a plan, their obligations to do their parts are mutually interdependent as much as their actions. Consider the obligations that would be created, on the other hand, if the generals planned to attack on the basis of a conditional promise followed by an unconditional one. If the general who made the conditional promise were to fail to follow through, the obligations of the general who made the unconditional promise would remain completely intact. If, on the other hand, the general who makes the unconditional promise were to fail to follow through then the obligations of the general who made the conditional promise would be null and void. For this reason, the current proposal fails to respect the sense in which plans create normatively symmetrical reasons for parts.[78]

There does not seem to be a good way for separatist to account for coordination in a way that respects the symmetry constraint. This constraint's force, then, accounts for the lack of explicitness in discussions of coordinated attack discussed earlier. Making the mechanism of coordination clear, assuming the separatist view, would make it clear that the target phenomenon had fallen out of view. The natural reaction to this problem is to obscure both the topic of coordination itself and the process by which agents bring it about. Of course, bullets can be

[77] Joe Mintoff (2004) has argued for a similar view to the one under consideration here: that agreements can be understood as an exchange of intentions. I do not have space to address this view directly, but I take it that an analogous objection to the one I raise here could be posed for Mintoff's view.

[78] I take this to be the most natural suggestion about how to construct plans out of conditional promises, but there are others. See Gilbert 1993 for a critical discussion of these other ways in a similar vein as mine.

bitten. One could stipulate that one is simply interested in whether the generals can attack simultaneously. This response, though, has a significant drawback. As we saw earlier, what gave rise to the puzzle was the palpable sense that the generals should be able to *coordinate* with one another. To give up on a meaningful sense of the term, then, is to give up on the problem we set out to solve.[79]

4. The Interpersonalist Alternative

The interpersonalist says that when agents form a plan through communication, their agreement is itself the source of each agent's entitlement to participate in the plan. This means the interpersonalist has an additional resource for overcoming the problem of rational interdependence not available to the separatist. The problem, recall, was that the grounds of each general's decision referenced the other's general decision. This circular structure– A's decision depending on B's and B's depending on A's– led to a stalemate. If the interpersonalist is right there exists a way out of the circle, one that respects the symmetry constraint. Each general can find themselves motivated to attack not by knowing what the other is going to do but by knowing that the two of them have agreed to a plan. The creation of a plan, then, can potentially be a fixed point– a fact that does not encode information about a prior decision of either general– around which the two can coordinate. This gives the interpersonalist a natural explanation of the agential and normative symmetry inherent in plans. For two parties, A and B, who make a plan, A relates to B's part in the same way that B relates to B's part, namely as something to be motivated by the communicative exchange that established their plan. The same point holds for how B relates to A. Furthermore, because A's and B's reasons have a common source the normative implications

[79] The caveat given earlier, that my argument can be reframed in terms of cooperation or plan-making, applies here as well. In fact, the symmetry constraint is even harder to deny as a constraint on cooperation or plan-making. For reasons discussed earlier and because I believe the symmetry constraint does apply to coordination, I frame my argument in terms of coordination.

of their neglect are also symmetrical. A's failing to heed A's reasons to do A's part undermines A and B's plan which directly undermines B's reasons for doing B's part.

Interpersonalism is a view about how agents become rationally entitled to perform certain coordinative behaviors, one that sets itself against a narrowly individualistic view of how such behaviors arise. It has similarities, therefore, with theories of team reasoning as developed by Sugden (2000, 2003, 2018), Bacharach (1999, 2006), and others.[80] For the purposes of understanding interpersonalism's place among existing views, then, we need to explore the extent of these similarities. Team Reasoning theorists were inspired by the observation that certain coordination problems– such as the Hi-Lo problem and Stag Hunt– are intuitively solvable but not within the confines of traditional game theory. As a remedy, they argued we can understand agents who identify as part of a team as being moved to act not in light of what is best for each of them individually but in light of what scheme of behavior is best for the team as a whole. So interpersonalism and team reasoning both contrast themselves with standard decision theory and seek to provide an alternative explanation of how coordination arises in certain instances. The important difference between them is the mechanisms of coordination that the two theories go on to posit and by extension the circumstances in which they see these mechanisms as operating. Interpersonalism is fundamentally a theory of communicative exchanges while team reasoning is a theory of a certain style of reasoning, team reasoning. Such team reasoning is neither necessary nor sufficient for the kind of interpersonal exchange that has been at issue here and is at stake in coordinated attack. On the one hand, agents can agree to a plan without identifying as part of a group, for instance when two corporations make a contract

[80] Further explorations of a team reasoning framework are made by Gold and Sugden (2007) and Colman and Gold (2017). Colombo and Guala (2023) provide a helpful overview of the literature.

with one another for purely self-interested reasons.[81] On the other, agents can reason as a team without communicating with one another.[82] Examples of this would be the very problems that initially motivated theories of team reasoning, such as Hi-Lo style problems, where no communication is possible.[83] The difference between interpersonalism and team reasoning theories, then, is evinced by the kinds of coordination problems each is trained on. Team reasoning theorists are interested in cases where agents are presented with a fixed set of possible actions and must do what is in their collective self-interest without being able to make a strategy explicit. Interpersonalism, on the other hand, is motivated by reflection on the coordinated attack problem, in which the agents' only hope to succeed is precisely to make such an arrangement. Indeed, not only do team reasoning theorists eschew communication as a paradigmatic means of effecting team behavior, there is no point that I am aware of where they offer a discussion of coordinated attack or similar problems where communication is at issue.[84]

[81] In this sense, the generals do not need to think of themselves as being on the same "team" or side in order to succeed. This might be a point of confusion about the coordinated attack problem since the generals are typically presented in this way, but such a characterization is not essential to the problem.

[82] See, for example, Sugden (2000) and Bacharach (2006). There are two motivations that, as far as I can tell, team reasoning theorists have for declining to make communication an essential aspect of their theories. First, they take the fact that agents can succeed in Hi-Lo style cases as evidence that team reasoning is possible without communication. Second, they view communication as itself just another form of coordinative behavior, and if that is right, a theory of coordination which appeals to communication is in danger of circularity. For instance, Sugden (2000, p. 180) anticipating the objection to his view that "most coordination is achieved through the use of language" responds that "language itself is a form of social coordination which needs to be explained, and which might plausibly be understood as emerging out of more basic tacit understandings". I have my doubts about both of these motivations but developing them would take us too far afield.

[83] For the reasons just given, I believe we should view interpersonalism and team reasoning theories as distinct but compatible amendments to standard decision theory. It should be noted, however, that there is the potential for a more adversarial relationship, depending on (a) the range of cases each treats as within its purview and (b) how unified a view of coordination each aspires to. The stakes of this debate and its prospects are complicated and fully explaining them goes beyond the scope of the paper. In the event of such a debate, my worry for team reasoning theorists would be whether they can give a satisfactory account of cases like coordinated attack in terms of a view initially designed for cases where no communication is possible. I believe an explicit agreement unequivocally rationally entitles agents to coordinate with one another. To fully make sense of such agreement, then, team reasoning theorists would need to provide not just an account of team reasoning but an account of what entitles agents to engage in team reasoning in the first place. I am skeptical that such an account is forthcoming without appeal to the primitive fact that communication enables coordination. In their overview of team reasoning, Gold and Colman (2020) similarly conclude that there may be "insuperable obstacles to solving the problem of rational group identification within a theory of purely instrumental rationality" (p. 314).

[84] Another example would be the electronic mail game.

The fact that the interpersonalist offers elegant solutions to the theoretical issues surrounding coordination suggests we might fruitfully reframe the coordinated attack problem in light of it. To do so we need to say more about the process of plan making itself. Earlier we asked how, in saying things like "let's meet tomorrow morning in my office" and "okay, sounds good", agents create a plan. The separatist adopts what could be described as an epistemic reading of this sort of exchange. The difference that performing a speech act such as "let's meet…" or "okay…" makes is thought to be that the intended hearer will simply acquire new information: the hearer then knows the act has been performed and, perhaps, knows it was done with a certain intention. On this picture, "accepting" and "proposing" bear no essential connection to the parties making a plan but one that is mediated by each individual's practical deliberations. The interpersonalist needs their own story about where reasons for parts come from. The obvious resource is the fact that agents almost always have a preexisting understanding between them of what it takes to make a plan, a plan making convention.[85] The interpersonalist can say that the satisfaction of such a convention itself suffices to give agents a plan.[86] When agents say things like "Let's do x" and "sure okay" the significance is the triggering of a certain convention's conditions whereby the agents, together, generate reasons to do their respective parts. Just as a speaker of English, in the right circumstances, can incur a promissory obligation by saying "I promise…" so, I submit, agents can give themselves reason to do their parts in a plan by saying "how about…" and "Okay".

Crucial for the purposes of getting clear on coordinated attack is noticing that the conditions of plan making conventions, perhaps more so than other convention governed

[85] I take it that insofar as some sort of mutual knowledge of convention like this would be required for plan making, it would also be required for performing communicative acts like asserting and thus that that in imagining the generals to have the ability to plan we are not altering the conditions of the original scenario.
[86] A notable recent defense of the idea that communication is essentially a matter of satisfying conventions can be found in Lepore and Stone 2015.

activities, are likely to be highly context sensitive. So far we have imagined plan making between two agents as having a typical propose-accept-structure: one agent makes a proposal– which they might do by saying such phrases as “how about", “let’s” or “I propose”– and the other agent accepts– which they might do by saying such phrases such as “okay”, “sounds good”, or, again, “I accept”. But not only are things more fine grained than this, there are also wholly other conventional structures.[87] Some situations may dictate that an act of confirmation on the part of the proposer is necessary to get a plan going, but in others it might be that agents have an understanding among them that a plan is created insofar as any agent in the group declares one. There might be situations where no formal acceptance is required and where any agent can simply establish a plan by putting one forward.

Another important axis of variability will be the channels of communication required by different conventions. Obviously for everyday social arrangements, it is often okay to convey acts like accepting and proposing through mediums like text message and email. In some contexts, simply leaving a physical message in a common area may be appropriate. At the same time there are clearly conventions which require the use of less noisy channels. When getting married, for instance, it is generally required that one physically be in the presence of one’s partner.

Finally, conventions may have conditions that are not obviously communicative in nature. When making a legal contract, for instance, the main condition is the creation of a certain physical object, a document bearing the participants signatures and submitted in a specific manner at a specific location. In certain deliberative contexts it might also be part of the plan making convention that the parties find themselves in a certain state of mind or with certain attitudes, independently of whether those states are expressed. For instance, in some contexts it

[87] See de Kenessey 2022 for an illuminating discussion of this topic.

might be a condition on making a plan that all agents genuinely want to be part of the plan. Finally we even could imagine a convention where the parties, in order to satisfy it, must *first* individually form intentions to do their parts. This would be an odd, somewhat redundant, convention, but it is one that starts to look strikingly like the separatist picture of plan making. And perhaps in exceptional circumstances such a convention is reasonable i.e. in situations where it is important that agents have reasons to do their parts reinforced in multiple ways.[88]

The question of what convention exists in a given group and by extension what communicative channels are open to its members is simply that, a question about that particular group and not one that can be answered in the abstract. One presumably important factor will be the costs associated with confusion. There is always the possibility, as we will see, that agents get their wires crossed as to whether a plan has been created, and the ramifications of such confusion will make different conventions more or less appropriate. When making normal social plans with friends, the cost of such a mix-up is rarely very significant. This means that the channel for expressing proposal and acceptance can be something like texting or emailing. For agents establishing a complicated legal relationship the cost of a mix-up may be very great, making more onerous conventions prudent. Also relevant will be the costs typically associated with inaction or failing to make a plan. The important thing is that different conventions will have different satisfaction conditions, which may depend on using a specific communicative channel, and that a plan is realized if and only if such a convention is satisfied.

5. Back to Coordinated Attack

The crucial insight that the interpersonalist has to offer the debate on coordinated attack is that we have principled way of denying one of the main assumptions made by the classical

[88] For a discussion of how individual and shared intentions to participate in joint actions might mutually reinforce one another see Alonso 2024.

account, namely that in order to have reason to attack a given general would have to first know that the other general was going to attack at the same time. Contra separatistism, to enter a plan, parties do not require foreknowledge of what the other parties are going to do. Thus, the generals do not need to establish common knowledge in order to be justified in attacking. Rather, plans themselves give agents reasons to do parts, and the generals can find themselves entitled to attack simply by forming a plan. For the generals, this could be as simple as one general sending a message proposing a time of attack and the other sending a message in return accepting that proposal. The messages could read: “let’s attack at dawn” and “okay”, presuming they have a propose-accept plan-making convention. In such a scenario the generals, having satisfied the relevant convention, would be fully entitled to attack at the agreed upon time.The same thing could be accomplished by one general sending a single message declaring a plan, assuming that the generals have the mutual understanding that this sort of unilateral plan making is appropriate. Solving the coordinated attack problem, in this sense, does not have to do with rethinking the amount of certainty the generals need in order to coordinate but with rethinking the content of the beliefs they need to form. If we assume that what the generals need to form beliefs about is one another’s future behavior, the generals find themselves in a vicious cycle from which no escape is possible. If we realize that what the generals need to form beliefs about is the extent to which a certain convention is satisfied no such circle exists.

By the lights of the interpersonalist, then, there are no in principle barriers to the generals coordinating. It just depends on what convention the generals actually have between them. On this point coordinated attack is admittedly under described. One could ask what convention it would be reasonable for them to have, but, this question, I take it, would connect with issues that go beyond the initial thought experiment. It would depend on what linguistic community the generals were part of, what situations that community typically confronted– the typical costs to confusion or inaction– and its actual socio-linguistic history. In other words it is a question for an expert in military strategy rather than decision theory. Nonetheless the interpersonalist is in a position to say that both that coordination is a live possibility and that it would be realized if the generals sent a series of messages back and forth confirming a time of attack. This is just the intuition we set out to vindicate.

One might worry, at this point, that the problem described by the classical account returns even on the interpersonalist picture. Recall, the initial roadblock to coordination was that the generals, not knowing their messages were delivered, are not in a position to ever know an attack is called for. Even on the view that the generals need to make a plan by satisfying a plan-making convention, still the generals need to know they have a plan and thus, one might think, they need to know that their messages are being delivered. There's three important issues here that should be taken in sequence.

First: there's the thought that in order to satisfy a plan making convention– say accepting a proposal– both generals need to know it is being satisfied. That is, you might think that the generals without knowing their messages are being delivered are not in a position to perform plan making gestures at all, even while granting the interpersonalist position. One might be moved by the thought that there is a conceptual connection between convention condition satisfaction and the relevant agents' knowledge of their satisfaction, which makes it look as though the sender of the acceptance needs to know that their message is delivered. Here is my response. It is true, generically speaking, that convention conditions are known to be satisfied, but we should not confuse this with thinking that this is a necessary condition of their satisfaction. Consider a ref blowing the whistle to call a play dead. Whistles are made to be heard, and thus it is no accident that the players ordinarily hear the whistle blow. But this does not mean that play does not stop if the players do not hear the whistle. Play stops when the whistle is blown. In the defective case where the players fail to hear it, they are in a sense making a mistake, perhaps an excusable one, in carrying on as though the play is not dead, but they are doing just that. This suggests that it is not true that convention conditions need to be known to be satisfied in order to be satisfied, indeed building in such a knowledge condition would seem to result in a regress problem: if a convention is not satisfied unless all the relevant parties know that it is satisfied then it would seem to be a further condition that the agents know that they know this and so on and so forth.

Second, there's the thought that there is a conceptual connection between having a plan and knowledge of that plan. I am sympathetic with this thought as well. The function of a plan is to coordinate agents, and a plan cannot do this work if the parties to it are not aware of it. This goes together with what we just said: typically plan making conventions are such that when they are satisfied agents know them to be satisfied. But we should not take this thought too far: plans are not necessarily known by the parties to it. Consider a case where a friend texts you proposing to meet at a certain time and you text back agreeing. Now imagine you run into them the day before and they say something suggesting they are not aware of your plan. The natural thing to do is not to reaccept their invitation but to simply inform them of the fact that you already have a plan with them. The fact that informing rather than "reaccepting" is the natural course of action suggests that in faulty cases where there has been some confusion it is possible for agents to have a plan without all of them knowing it.

Third, finally, there's the thought that the possibility of such confusion makes satisfying a plan making convention overly risky for the generals in particular. If one general sends a message of acceptance and, as it happens that message fails to be delivered, their convention might be such that they really do have a plan but one that the recipient simply does not know about. One might think that it is precisely because it is possible, if unusual, for agents to have a plan without knowing it that the generals, knowing this, would be irrational to attempt to make one. If that's right then we are back to square one: the generals cannot attack because their way of communicating is not reliable.

This worry underestimates the idea we encountered earlier: that conventions have baked into them the communicative conditions under which they are employable. If the generals have a convention that is operative between them, then their standing disposition will simply be to draw on it. Again, the likelihood of a given convention in fact being operative for the generals is more a question of military strategy than one whose answer can be read off of the generals' individual preferences. Indeed, there surely are situations where agents are constrained, in one way or another, by the fact that their lines of communication are too noisy, and we should not expect our best theories of communication to rule this out. The goal here has been to make sense of the intuition, *assuming one has it*, that coordination is possible. Once we see that the generals can be justified by satisfying a plan-making convention rather than by gaining knowledge of what the other will do, we see that coordination is indeed possible.

6. Conclusion

Entering into a plan is not as epistemically onerous as is often assumed. It can be as simple as successfully sending messages such as "let's attack at dawn" and "okay". Of course, much more remains to be said about how agents, just in saying such things, thereby make a plan. The intentions held by those agents, their setting and material conditions, the behavioral and institutional effects of their acts all bear a complicated metaphysical relationship to the conventions they employ.[89] To give an account of how these issues fit together would be nothing less than to give an account of how it is that speakers of a language, as Austin (1962) put it, do things with words. One thing to be acknowledged is that in order to avail themselves of a convention agents presumably need to know something about both it and one another. Indeed, it may be that a convention needs to be public among the agents that draw on it, and it may be that this sort of publicity requires common knowledge.[90] Even if that is right, though, the interpersonalist account suggests an altogether different connection between common knowledge and coordination than the one that is typically taken for granted. Rather than think of the generals as trying to coordinate by going from having no common knowledge to having common knowledge, we can think of the common knowledge they inevitably already have as making possible a shared practical achievement: coordinating through plans.

[89] For interesting recent discussions of these issues see Unnsteinsson (2019), Kukla (2022), and Wanderer (forthcoming).

[90] For an illuminating discussion of the connection between publicity and common knowledge see Lederman (2018).

3. The Interpersonal View of Testimony

1. Introduction

Philosophers use the term *testimony* in an idiosyncratic way. By testimony they do not mean eyewitness accounts offered in court room settings or stories of religious transformation but simply any occasion on which one person informs another of some fact. You provide testimony, on this way of using the term, any time you perform an ordinary act of telling someone something.

There are a number of philosophical perspectives from which one can approach this topic. One common way of exciting interest in it is to point out the enormous extent to which we rely on what other people tell us. A huge number of our beliefs are based on the word of others. These include not just trivial or isolated beliefs but also beliefs which play foundational roles in our worldviews: beliefs about the course of human history, about locations that we have never been to, about the nature of the physical world, about our own personal history, and so on. This observation may induce a kind of epistemological vertigo: if testimony is not reliable, then one's whole belief system is on shaky footing and a dramatic reorientation would need to take place. It thus becomes a pressing philosophical task to establish that another's word can legitimately ground knowledge.

We can also begin to wonder about testimony by considering the perspective of its producer rather than of its consumer. It is common when reflecting on practices like promising or consenting to marvel at the fact that we can alter the normative landscape of those in earshot simply by saying "I promise to…" or "I consent to…". But it is no less striking that one can alter the epistemic situation of persons nearby by merely uttering a fact. How is such an other-facing epistemic power so much as possible, we might ask.

A final illustration of the philosophical importance of testimony points to the degree to which testimony permeates our social lives. While it might be going too far to say that all meaningful human relationships depend on testimony, our ability to tell one another things and be believed plays a foundational role in a wide variety of relationships and social interactions. It is in virtue of being able to communicate with others that I am able to coordinate my actions with theirs. Furthermore, the ability to tell others about my life and self is essential to being understood by them and therefore to my most important social connections.

How one motivates the topic of testimony is likely to influence the outcome of one's inquiry. For instance, an inquiry into testimony from this last point of view might not have as its ultimate goal an understanding of how testimony produces knowledge, although that will be a guiding consideration. Instead, its aim will be to provide a satisfactory account of what exactly the nature of the relationship is between speaker and addressee in a case of testimony and how the interaction that transpires between the two makes possible a certain kind of collective agency.

However, I believe each of these perspectives points to an important truth about testimony: testimony is produced with a certain intention or aim; it creates a distinctive relationship between its producer and its intended audience; and this relationship has a transformative effect on the audience's epistemic situation. The deeper challenge of understanding testimony, I suggest, rests in articulating these truths together, in a unified way. In this chapter, I want to lay the groundwork for my proposal for meeting this challenge by exploring the prospects of a family of views in the literature, which stress testimony's intrinsically social nature. Specifically, I am interested in views which treat the testimonial relationship between speaker and hearer as the primary source of epistemic significance in a testimonial exchange. Following others, I call such views "interpersonal views".[91] After arguing for the prima facie attractiveness of interpersonalism, I outline two questions for this kind of view, questions which I believe have not been adequately addressed. The first is how such views account for seemingly impersonal factors such as the speaker's reliability, sincerity, and competence.[92] The second is how such views make good on their claim that there exist special interpersonal or "non-evidential" reasons for belief. I conclude that finding answers to these questions remains a pressing challenge for those committed to the intuitions driving interpersonalism.

2. Interpersonalism

There are many debates in the literature on testimony– the question of whether testimony can generate knowledge in addition to transmitting it[93]; the question of whether testimony should be thought of evidence[94]; the question of whether groups, in addition to individuals, can provide testimony[95]. However, by far, the central and quintessential debate in the literature is between reductionists and non-reductionists.[96]

Reductionists claim that testimony is not a self-standing or fundamental source of knowledge. Rather, according to reductionism, testimony-based beliefs are products of more

---

[91] Interpersonalist views have been offered by Ross (1986), Hinchman (2005), Moran (2018), Faulkner (2007), and McMyler (2011)

[92] Specifically, I focus on an influential version of this question posed by Lackey (2008).

[93] In particular, Lackey (2008) is known for defending the view that testimony can generate knowledge. Faulkner (2011) has explicitly defended the thesis that knowledge can only be transmitted. For a recent discussion see Leonard (2018).

[94] I will take up this question in a later section. The position that testimony is not evidence usually overlaps with interpersonalist views of testimony. The claim that testimony should be thought of as evidence is explicitly defended by Leonard (2016).

[95] See Tollefsen (2007) and Lackey (2014) for discussions of how to think about testimony provided by a group.

[96] Some prefer "anti-reductionist" to "non-reductionist". While the term "anti-reductionist" is often accurate, it is more narrow than "non-reductionist" so I prefer "non-reductionism" as a catch-all term. Reductionism traces back to the work of David Hume (1740). More recently, reductionism has been defended by Fricker (1987) and Malmgren (2006), for instance. Non-reductionism can be traced back to the work of Thomas Reid (2002) and more recently by Coady (1992) and McMyler (2011).

familiar kinds of belief-forming processes. So, for example, one might think that when a hearer believes *p* on the basis of a speaker saying *p*, what happens is: the hearer *perceives* the speaker saying p, the hearer *remembers* that the speaker has spoken truly in the past, and the hearer, therefore, *infers* that p is true. Thus, we seem to have a plausible explanation for how a hearer could form a legitimate belief in a speaker's testimony, the mechanics of which simply appeal to independently understood sources of knowledge: perception, memory, and inference. Reductionists seek an account like this, one which reduces testimonial justification to non-testimonial justification.

Non-reductionists reject such accounts, seeing in testimony a basic source of justification. Of course, non-reductionists acknowledge that a speaker's testimonial track record can undermine the legitimacy of their testimony. They maintain, however, that normally, in the absence of such defeaters testimony itself provides a hearer with a reason to believe the proposition testified to that is sufficient and sui generis.

The pull towards non-reductionism comes from reflection on the psychology or phenomenology of the two poles of testimony: speaker and hearer. Consider the hearer. Typically, when one receives testimony one accepts it immediately. Considerations of the speaker's reliability or the likelihood of their speaking truly are not typically present in one's mind. For this reason, the reductionist view seems to over intellectualize the process of receiving testimony. A corresponding point holds for the speaker's state of mind when testifying. When one tells another some fact, one typically conceives of oneself as offering a conclusive reason to believe the fact testified to. One intends not merely to add more grist to the hearer's epistemic mill, but to settle the question for the hearer in a dispositive fashion.

The reductionist of course has resources for responding to these concerns of the non-reductionist. Frequently, reductionists argue that acts of memory, induction, and inference operate automatically, in the background of the hearer's mind. Moreover, the observations motivating the non-reductionist only characterize typical or paradigmatic testimonial exchanges, and for this reason are not yet decisive. Nonetheless, at first pass, considerations of the spirit in which testimony is typically offered and received provide powerful incentive to explore the viability of non-reductionist programs.

The interpersonalist approach to testimony is a natural elaboration of the non-reductionist approach. As we have said, interpersonalist views of testimony pay special attention to the distinctive relationship that testimony creates between speaker and hearer. Let us return to the observation that testimony tends to be offered as a conclusive epistemic reason rather than as material meant to enable the hearer to make up their own mind. It is interesting to note that the same point does *not* seem to hold for overhearers of testimony. If I merely overhear a speaker's testimony it is not obviously a violation of the spirit in which it was offered for me to consider their testimony as just one reason among many weighing in favor of the proposition testified to. The asymmetries between hearer and overhearer do not end there. It is often pointed out by interpersonalists that only the intended hearer of a piece of testimony thereby obtains certain rights and entitlements with respect to the speaker. For instance, an intended hearer has a right to complain if the speaker's testimony proves false and can defer to the speaker when pressed for evidence by others. Such asymmetries (between hearer and overhearer) are mutually suggestive. There seems to be a special testimonial relationship that obtains between a speaker and hearer. The interpersonalist takes this special relationship as her starting point for an account of testimony as a distinctive source of knowledge.

Different interpersonalists, of course, use different language to account for how testimony creates this relationship between speaker and hearer. Richard Moran (2005), for instance, stresses the fact that the speaker adopts certain responsibilities towards the hearer, describing testifying as providing “assurance” to a hearer . Ross (1986), on the other hand, describes a speaker as providing a hearer with a guarantee of the truth of the testified to proposition. Finally, Hinchman (2005) characterizes testimony as an “invitation to trust”. What unites these views is that they treat the normative relationship that binds speaker to hearer as the foundation of testimony’s epistemic significance rather than a concomitant phenomenon.

3. Lackey’s Objection

Testimony can obviously have epistemic significance. If one person asks another how they know something, a valid answer is that they were told it. At the same time, we have just seen that testimony can impact the normative relations between persons. Testimony confers certain rights and entitlements specifically to the intended audience. The interpersonalist takes an ambitious approach to the topic in that the interpersonalist seeks to understand these two aspects as a unified whole. However, there are a number of theoretical issues with interpersonalism concerning the feasibility of such a unification.

One well-known objection, in the form of a dilemma, has been put by Jennifer Lackey (2008, chapter 8). The two horns of the dilemma are constituted by two ways in which the interpersonalist might conceive of the epistemic significance to the hearer of the speaker’s reliability.

Following along one horn, background considerations of the speaker’s reliability, in a given testimonial exchange, simply do not have epistemic significance to the hearer. An interpersonalist might argue, in other words, that a hearer is entitled to accept a speaker’s testimony merely because the speaker provides their assurance or guarantee that the testified to proposition is true, and that such an entitlement does not depend on the hearer knowing anything about the speaker's reliability in general.

The problem, if the interpersonalist goes this way, is that their view becomes too epistemically permissive. It simply is the case that plenty of testifiers are either insincere, incompetent, or otherwise unreliable. Confronted with this fact, a competent epistemic agent should *not* accept just any testimony. But the interpersonalist seems to have no way of explaining why this is so. Instead, the interpersonalist is forced to say that such unreliable speakers *do* give hearers valid reasons for belief. This would be an unacceptable consequence of the interpersonalist view.

The second approach the interpersonalist might take in response to the dilemma is to admit that background considerations of reliability, sincerity, and competence *do* have some epistemic role to play in determining when a speaker provides a hearer with a legitimate, testimonial reason for belief. The problem facing the interpersonalist if they go down this route is that, however they spell out the way in which these considerations enter into a hearer's theoretical deliberation, there seems to be no epistemic work left for the interpersonal aspect of the testimonial exchange to perform. If a hearer is inclined to believe a speaker because the hearer deems them credible, reliable, and sincere then isn't the epistemic value of the speaker's testimony entirely determined by such factors? Traveling along the second horn of the dilemma, then, the interpersonalist fails to live up to their name: they fail to make good on their idea that the interpersonal transaction between speaker and hearer (the speaker's assurance, guarantee, what have you) itself has some sort of special epistemic significance.

Here is Lackey, to sum up:

> …either the view of testimony in question is genuinely interpersonal but not epistemological, or it is genuinely epistemological but not interpersonal. In the former case, the [interpersonal view] will be novel but useless for an epistemology of testimony; in the latter case, the [interpersonal view] will be epistemologically useful, but not interestingly different from its so-called competitors. The bottom line is this: interpersonal features cannot create justification or warrant epistemologically *ex nihilo* and hence there is no room for a genuinely *interpersonal* view in the epistemology of testimony. (2008, 240)

There have been recent attempts to respond to Lackey's objection. Judith Baker and Philip Clark (2018) argue that interpersonalism can be saved by appeal to the notion of epistemic buck passing. Their idea is that a speaker's testimony entitles a hearer to a belief in the testified to proposition that does not need to be justified by evidence the way an ordinary belief would. Rather, testimony makes it the case that a hearer can acknowledge to others that they do not have evidence for the testimonial-based belief when challenged to produce evidence and to defer to the speaker instead. On their view, then there is no special interpersonal justification for testimony. Rather, the testimony, when offered by a credible speaker, itself removes the need for such reasons via epistemic buck passing. We still get an interpersonal view, however, since the relationship between speaker and hearer does play an ineliminable role in entitling the hearer to a belief in what the speaker says.

I believe that Baker and Clark are right to respond to Lackey's dilemma by trying to come up with a view that acknowledges the importance of both non-interpersonal and interpersonal factors i.e. one that travels the second horn of the dilemma. However, I do not find their appeal to buck passing satisfying. Epistemic buck passing is something a hearer can do in response to third party demands for justification. But it is not clear to me that one can read off an agent's reason for belief by appealing to their entitlements with regards to such demands. Indeed, a hearer who accepts a speaker's testimony experiences herself as having a positive, valid reason for her new belief: the reason just is the speaker's testimony or so it seems to me. This brings me to a second worry, namely with the idea of epistemic buck passing itself. If someone asks me why I believe something, a perfectly adequate response is that someone else told me. Such a response does not strike me as deferring or passing the epistemic buck but simply stating one of the legitimate reasons one might have for belief. So at the very least Baker and Clark's solution requires appeal to a notion, which itself requires considerable further clarification.

Another response to Lackey is offered by Florencia Rimoldi and Federico Penelas (2025). As I understand them, Rimoldi and Penelas propose a version of interpersonalism on which a hearer has two kinds of reasons to believe in a speaker's testimony: (1) they are privy to the set of facts that an overhearer is privy to and therefore have impersonal grounds for believing what they are told and (2) they have a special, additional interpersonal reason for belief in virtue of being the addressee. Appropriately, they call this position "interpersonalism+": the interpersonal reason is added to a set of impersonal reasons. Furthermore, they claim that this additional interpersonal element, at least sometimes, still has epistemic "work" to do, for instance in situations where the stakes of the truth of some proposition is especially high.

I want to set aside the issues related to pragmatic encroachment that this view raises since I believe there is a deeper and more instructive issue at play. Rimoldi and Penelas' approach relies on dividing up the total amount of epistemic justification a hearer has as a result of receiving testimony. In doing so, they succeed in making a distinction between the epistemic position of a hearer of testimony and an overhearer. However, on their view, the lion's share of a hearer's epistemic justification for believing what they are told is still of the same kind as an overhearer's. Thus, as it were, they wind up with a view that is interpersonalist in letter but not in spirit. The problem, then, with their view points to the depth of Lackey's objection. Not just any story that involves considerations like a speaker's reliability and the speaker and hearer's relationship will do for the interpersonalist. Rather, considerations of sincerity and reliability themselves need to matter in an interpersonal way. The positive view I propose in the next chapter proceeds along such lines.

4. The Question of Evidence

Lackey's objection prompts the interpersonalist to say something about the epistemic significance to the hearer of arguably impersonal considerations like the speaker's reliability, competence, and sincerity. The second issue that I want to suggest is pressing for the interpersonalist concerns how they talk about the epistemic value of the testimonial relationship itself. In particular, interpersonalists, as a way of expressing the special interpersonal nature of testimony, are united around the claim that the reason generated by a speaker and hearer's relationship should not be thought of as *evidence* for the hearer.[97]

One way of starting to get a handle on what the interpersonalist means by this denial is by recalling a thought put forward by Grice (1957): that there are two senses in which the word "mean" can be used. On the one hand there's the way that a doctor, for instance, might use it, as in, "those spots mean measles" where "spots meaning" measles maps onto the idea of the spots being evidence for measles. On the other hand, we can also use "mean" as in: "those three rings of the bell mean that 'the bus is full'". Both the spots and rings of the bell, by meaning what they do, lead an agent to form a related belief, but they do so in different ways. In the case where the

[97] This position is what I will call non-evidentialism. It has been argued for in some form or another by Moran (2018), McMyler (2011), and Hinchman (2005) as well as others.

bells mean ‘the bus is full’ a whole proposition is embedded in the second half of the means-expression. Like instances of testimony, the rings of the bell have expressive content. The spots, on the other hand, simply mean measles not “there is measles”.

Intriguing as it is, the position that testimony is not evidence has been met with stiff resistance and become one of the key disputes spurred by interpersonal accounts of testimony.[98] For one thing, it is not clear that the Gricean distinction reproduced above actually maps onto the way the term evidence is used by most epistemologists who are thinking about testimony. The more common understanding is simply that evidence is that which justifies belief.[99] And so far the Gricean distinction does not show that testimony fails to be evidence in this broad sense. This failure gives the impression that what there is to say about testimony’s social importance is after all to the side of the right account of its ability to give an individual agent a valid reason for belief.

To make progress, I turn to the work of Richard Moran (2018), one of the philosophers most associated with the view that testimony is not evidence. For Moran, one sees the truth of non-evidentialism from an appreciation of the metaphysical nature of the testimonial act, which involves presenting oneself to the addressee as a guarantor of the proposition.[100] This view gets unfolded by considering each side of the testimonial act. On the one hand, Moran clarifies the specific intention with which testimony is offered and on the other the specific way in which the addressee, in accepting that testimony, responds to that intention. Taken together these roles place constraints on the kind of reason that can be produced, giving it, according to him, a non-evidential character.

---

[98] See, for instance, Vermaire (2020) and Leonard (2016) for critical discussions of the thesis that testimony is not evidence.
[99] See Kim (1988) and Fricker (2019) for defenses of using the term “evidence” this way.
[100] Moran, 2018, 45

In particular, Moran follows Grice in noticing the self-reflexive character of the intention with which testimony is offered by a speaker. Not only does the speaker seek to generate a belief in the addressee, they seek to do so by making manifest to the addressee that very intention. Thus, for the testimonial act to be completed the addressee and speaker must stand in a relation of mutual recognition with the speaker's intention. The recognition at issue here is not mere theoretical one– an observation of the speaker's intention– but a practical endorsement of the speaker's authority to give such a reason to the addressee at will. The speaker and addressee mutually acknowledge not just that, in an objective sense a reason has been offered, but that a valid reason has been offered. For Moran the special relationship to the speaker's intention the testimonial parties have secures the conclusion that testimony cannot be evidence:

> "Telling aims at being believed, which proceeds, via the speaker's overt assumption of responsibility, by joining together the particular belief proposed for acceptance, the kind of reason being presented for it, and the reason accepted by the audience. An evidential stance, by contrast, decouples all of these from each other, to be reassembled as the observer thinks best. But such a stance is contrary to the speaker's perspective on her act, in so far as it pictures her presentation of herself as meaning, in effect, that as far as reason-giving force goes, the audience is on his own; as if the meaning of her utterance were, 'Now I have spoken, make of it what you will,' rather than, 'Take it from me.'" (2018, 74).

In this passage Moran makes an appealing suggestion about how to understand the difference between testimony in *p* being taken in an evidential and a non-evidential way: that it corresponds to a difference in attitude the addressee can have, between being "on their own" with respect to whether *p* and not being "on their own". As a way of drawing out what is right about this distinction we can imagine a solitary epistemic agent reflecting on whether some proposition is true. Making up their mind will involve bringing to bear everything that they know about the world, weighing all the facts that they are privy to, and seeing in which direction they point, whether to the truth of the proposition being inquired into or its falsity. Such an agent, we would say, should consider all the evidence that they have and see where that evidence points. Now imagine that this agent is interrupted by another person and is directly told by that person that the proposition they are inquiring into is true. And imagine our inquirer takes the interrupter's word for it. A way of hearing Moran's thought in the above passage is that, in this imaginary scenario, the interrupter's testimony is not some new information to be folded into the original inquiry. Instead we can think of the speaker as advising the inquirer so as to make the inquiry unnecessary, an effect that only another agent could hope to have. The distinction between evidence and testimony, therefore, looks like a distinction in kind between the raw material of a theoretical inquiry and meta-inquiry advice about how to conduct it or when to close it.

This distinction is intuitive, but it does not get us all the way to the idea that testimony is fundamentally different from all forms of evidence. A reasoner's evidence, one might think, does not just refer to the information available to them when they are actively involved in an inquiry. Crucially, it seems that one may receive testimony at a point when one has not been conducting an inquiry into the proposition mentioned. One may be told something about an issue that one has never thought of before. So we cannot capture what is distinctive about testimonial reasons in general by relating testimony to pre-existing inquiries e.g. as advice about how to conduct them. Moreover, while Moran mentions an evidential "stance," we might be able to replace the word "evidential" with something else– a "neutral" or "interrogative" stance– while leaving the original point intact. If so, Moran may have something to teach us about the perspective testimony demands an addressee to take without teaching us something about whether testimony is evidence.

To adjudicate this issue we need to secure a definition of evidence that allows us to say whether testimony falls under it or not. While Moran does not provide one he is clear that as far as he is considered, the paradigmatic evidential setting is a forensic one in which an article of evidence speaks in favor of some proposition independent of the intention with which it is presented.[101] Consider how you might react to someone showing you a photograph meant to demonstrate some fact (that so-and-so was at the party, for instance). In such a case the reason that you get from looking at the photograph is separable from the intention with which the person shows it to you. The photograph, the evidence, stands on its own, as something with epistemic force for you, the viewer, the same as it would if you stumbled upon it accidentally. So when Moran uses the term evidence he has in mind this kind of thing: a physical article or object causally connected to the fact that it is evidence for. A footprint is evidence for a footfall; a bloody glove for a murder. The epistemic force of a speaker's utterances, on the other hand, is bound up with the intention with which the utterance is presented.[102]

Moran's argument is this: evidence's epistemic force is independent of the intention with which it is presented; utterances do not have any epistemic force independent of the intention with which they are presented; therefore testimony is not evidence. The problem with this form of non-evidentialism is that it relies on a hidden and obviously false premise that utterances (taken as mere sounds) are testimony. One of Moran's key points, however, is that testimony *as a whole* essentially involves the speaker's intention. If that is so then it is no surprise that the part of testimony that can be separated out from the speaker's intention cannot be considered

[101] See especially Moran 2018 pp. 50-52.
[102] In a reply to Fricker, Moran (2019) confirms that the definition of evidence he has in mind is idiosyncratic in that it is not coextensive with "truth-related considerations" or "epistemic reasons."

evidence on its own. Moran's thought about how testimony makes a certain perspective available to an addressee (of no longer being on their own) was a thought not about utterances but testimony as a whole.

Since we cannot do justice to Moran's intriguing thought by separating testimony and evidence as he uses the term, it is fortunate that the impulse to think of testimony as evidence is rooted in a much more permissive understanding than Moran's anyway. As we have said, the term "evidence" standardly serves to pick out anything and everything that justifies an individual agent's beliefs.[103] "Evidence" in this sense includes much more than the kinds of things one might find at a crime scene and would include memories, perceptions, mathematical intuitions, and so on. This permissive view of evidence is so taken for granted that it is hard to find specific advocates of it. Elizabeth Fricker though helpfully voices it in responding to Moran. Here's Fricker:

> Evidence and only evidence does and ever can serve as a justifier for belief; where evidence is understood to be whatever satisfies this principle: E is for S evidence that P relative to her current belief set $B_{1....}B_n$ just if adding E to $B_{1....}B_n$ increases S's subjective probability that P: P is more likely to be true in her epistemic position post receiving E, than in her epistemic position prior to receiving E. (2019, 769)

For Fricker to say that testimony is evidence is to say that the fact that someone has testified to P can be for the hearer something which raises their subjective probability that P. So Fricker, in thinking that testimony is evidence, need not think of the literal sounds coming out of the speaker's mouth as physical evidence that is manufactured with a separable communicative intention. Rather the fact of the intentional speech act taken as an avowedly unified whole is the relevant fact, the evidence. There's room then to respect Moran's point about the self-reflexive

[103] Again, see Kim (1988) and Fricker (2019) for examples of presentations of this understanding of the term.

structure of testimonial intentions while treating testimony as evidence in Fricker's broader sense.

Fricker's definition of evidence, in terms of subjective probabilities, is just one way of voicing a more general commitment to evidence being understood as that which justifies belief. By calling it into question, at least by name, proponents of non-evidentialism invite us to reflect on the status of what appears to be an innocuous bit of terminology and an unavoidable consequence of it. "Evidence" by definition is what justifies belief, testimony justifies belief, ergo testimony is evidence.

Recall that the impulse behind non-evidentialism was to do justice to the interpersonal aspect of testimony, to capture the distinctive way in which a speaker can make an epistemic difference in the mind of another agent. If this aspect is in tension with evidentialism it must be because the terms evidentialism treats as basic– namely, individual agents who have certain beliefs which are more or less justified– do not constitute a vocabulary rich enough to express it. Perhaps then the thought of the non-evidentialist, extrapolated to a position which opposes what is in fact evidentialism, is that testimony is productive of a state that is genuinely epistemic or world-disclosing but which cannot be understood as straightforwardly belonging to one individual or another.

There is also a metaphysical dimension to the question under discussion here that needs to be unpacked. Obviously, the claim that testimony is not evidence is not that all instances of testimony are evidence for the propositions they correspond to. Speakers lie, lose credibility, and get things wrong all the time. But even if they did not, evidentialism would not follow. To gloss evidentialism as an empirical statement about actual instances of testimony would require us to already have in hand an established understanding of what testimony is, something that could license us to pick out individual instances of testimony and ask of them whether they are evidence for what they assert. The claim that testimony is not evidence, on the other hand, is meant as part of an account of what testimony is in the first place. So while it may be true that most instances of testimony are in fact evidence for their corresponding proposition, presumably there is a deeper point at stake: a point about what testimony fundamentally is.

A natural thought is that evidentialism is the claim that testimony that *p* is evidence for *p* "when all goes well"– just as facts about thermometers are evidence for the temperature when all goes well with thermometers. But this way of putting it is only helpful if we have some way of expanding on what it means for testimony to go well, of saying what it is that makes some conditions count as the felicitous ones for testimony and others not. We cannot simply say that the right conditions– the speaker's being sincere, reliable, and able to communicate what they mean effectively– are determined by whatever makes the testimony evidence. If we did, we would render the evidentialist's account tautological. Of course, everything is evidence under the conditions that make it evidence.

A way of making progress on clarifying the sense of testimony *being* one thing or another returns to the observation that ordinarily one provides testimony for a reason. One does it intentionally. When I tell my roommate we need coffee this is not done haphazardly but because I have some specific motivation, some reason to inform him of the situation. Testimony's practical context, the fact that it serves a specific need or function in the lives of the agents it relates, may explain why certain conditions count as good conditions for it. When the speaker is seen by the addressee as sincere and reliable these are the conditions that are conducive to the aims of the speaker or the addressee who has solicited the testimony. The evidentialist, therefore, in thinking testimony as such justifies belief makes a seemingly plausible assumption:

that testimony's practical context is any situation in which one person, the speaker, senses that another, the addressee, lacks a belief that the speaker possesses. By highlighting testimony's practical context we finally have a sense of what evidentialism amounts to: the view that the aim with which a speaker characteristically testifies is that of generating a belief in the addressee. For interpersonalism to succeed, then, we would need a strong argument against this claim; otherwise, we will have to abandon their central strategy for positing a special "interpersonal" testimonial reason.

5. Conclusion

The most natural development of non-reductionism with respect to testimony leads to an interpersonalist approach to the topic. Moved by the moral-psychological asymmetries between a hearer and an overhearer, the interpersonalist seeks to give an account on which there is a fundamental *epistemological* asymmetry between hearer and overhearer. However, such an account leads to difficult questions (a) about the way in which seemingly impersonal but clearly relevant facts like the speaker's reliability are epistemically relevant to a hearer and (b) about what exactly it means to posit a distinctly interpersonal or non-evidential reason for belief. Absent answers to these questions the prospects for interpersonalism begin to darken.

4. Testimony, Transparency, and Shared Knowledge

1. Introduction

When trying to forward philosophical understanding of a certain entity, one strategy is to ask whether that entity has a certain function and if so what it is.[104] Certain theories of the good life, for instance, begin by considering what the characteristic activity of human beings is[105]; within philosophy of mind, functionalists explicate mental states by outlining the role they play in the system of which they are a part[106]; and one approach to social institutions begins by asking what good or social benefit a given institution is meant to bring about.

What about testimony? It seems clear that testimony, even more so than, say, a mental state or a human being, does have a particular function. This is arguably what Elizabeth Anscombe was on to when she famously said, "it is an insult and it may be an injury not to be believed".[107] If a speaker is not believed this is not just another fact about their testimonial act, such as the pitch of voice in which it was spoken or whether it was overheard by someone other than the addressee. Rather, a speaker not being believed means their act failed *qua* testimony.[108] Testimony, therefore, can be contrasted with other assertions of fact which merely serve to express a speaker's belief.[109] For example, I might say aloud to myself "I'm such a fool" without hoping to accomplish anything in particular or someone else might say to me, "you're such a

---

[104] Something's function can be understood as its typical or paradigmatic effect.
[105] Aristotle is associated with this sort of theory. For a discussion of his work along these lines see Korsgaard 2008 chapter 4.
[106] See, for example, Hillary Putnam 1960.
[107] Anscombe 1979 pp. 150. For a recent discussion see Marušić 2024.
[108] An assumption underlying this claim is that testimony has just one aim. One might disagree, thinking that testimony may have different aims that are operative depending on context and perhaps even different modes of uptake. I believe the assumption of uniformity is valid, but I won't defend it here.
[109] An interesting question raised by this contrast is whether there is a conceptual priority between these two kinds of speech acts. When one gives testimony does one express one's belief with a special testimonial intention or should we understand mere expressions of belief as instances of testimony which lack the normal testimonial aim? For a good discussion of this issue see Hinchman 2020.

fool" without trying to convince me.[110] Of course, a speaker often will have further ends in mind when they testify to some fact. A speaker may, for instance, intend to amuse their hearer by informing them of what they saw that day. But in that case, their end of amusing would not be had simply in virtue of being a testifer. When testimony is given there is always a distinctive testimonial aim in virtue of which we can ask: was their telling *qua* telling successful? [111]

It is surprising that epistemologists have not paid much attention to the aim of testimony.[112] There's at least two possible explanations for this absence. One explanation is that there is a nearby topic, assertion, which is heavily informed by discussion of its aim or function. It has been directly argued for and generally accepted that the aim of assertion is to add to the common ground of conversation[113]. Moreover, much work on assertion explicitly takes this consensus as a starting point. It is possible, then, that this neighboring literature acts as a kind of lightning rod redirecting discussion of the aim of testimony away from its natural outlet.

A second possible explanation for why epistemologists do not reflect much on the aim of testimony is that it is already obvious to them what it is: the aim of testimony is the intended hearer of that testimony believing the proposition testified to. Indeed, much of the discussion around testimony can be seen as a debate over what sort of reason for belief testimony provides a hearer. It is natural then to suppose that the various sides to this debate agree that the

[110] An interesting question raised by this contrast is whether there is a conceptual priority between these two kinds of speech acts. When one gives testimony does one express one's belief with a special testimonial intention or should we understand mere expressions of belief as instances of testimony which lack the normal testimonial aim? For a good discussion of this issue see Hinchman (2020).
[111] It's important to note that tellings can fail in other ways: a speaker can fail to be heard but also can fail to be understood as telling, for instance if the addressee takes the speaker to be performing a tongue twister. There is, therefore, a sense in which I have told you something just insofar as you understand what I am doing even if you end up disagreeing with me. But the mark of a telling being completely successful or fully consummated must be the addressee believing the speaker since understanding a speaker as telling is important precisely because it is a necessary condition on acceptance.
[112] Two exceptions are Coady (1992), whom I discussed later, and Lackey (2008).
[113] The literature on this issue is heavily influenced by the work of Stalnaker (1978, 2002).

characteristic aim of testimony is to provide such a reason. So perhaps, epistemologists do not discuss testimony's aim because their discussion already treats that question as settled.

In any event, it is a natural first thought that testimony aims at producing a belief or a set of beliefs related to the proposition testified to in its audience, that when my telling you *p* is fully completed, you simply believe *p*. I will call any belief-based view like this a *doxastic* view of the aim of testimony or a form of *doxasticism*. Doxasticism is suggested by the conjunction of two relatively uncontroversial starting points: first, that testimony is fundamentally an epistemic act and, second, that the only epistemically valuable state is belief. Together these ideas point to a doxastic view: testimony is about a speaker and addressee being cognizant of the world, and belief is that attitude in virtue of which an agent is cognizant of the world ergo testimony is about beliefs.

The goal of this chapter is threefold: first, to argue against doxasticism; second, to provide an alternative to it; and third, to use that alternative as the foundation for a version of interpersonalism that meets the challenges raised in the previous chapter. The problem with doxasticism emerges from reflection on cases where, intuitively, it is reasonable not only for someone to want to have certain beliefs but to be told something by someone in particular. I can care not just that I form a set of beliefs on some matter but that someone in particular tells me about it. Not every case of testimony is like this, but that such cases do exist points to an important feature of the aim of testimony in general. After outlining and motivating the two most appealing versions of doxasticism, I argue that any such view will struggle to account for the significance that address has in cases where the telling itself (by a particular speaker to a particular addressee) is valuable. After responding to a number of objections to my objection to doxasticism, I conclude that we need to rethink one of the supposedly uncontroversial ideas mentioned earlier. In order to explain the significance of address we find ourselves in need of a concept that is indeed epistemic– it is belief-like– but it is such as to be held or shared by multiple individuals. I flesh out the notion of a shared belief or shared knowledge by drawing on transparency accounts of self-knowledge and by discussing the role shared knowledge plays in joint action. Finally, I develop an epistemological account of testimony based on the idea that its aim is shared knowledge in the sense discussed in the previous sections. I show how this account constitutes a robust version of the interpersonalist position, one that adequately responds to the issues raised in the previous chapter.

2. Two Versions of Doxasticism

The most straightforward view about the aim of testimony is what I'll call *the simple doxastic view*: that testimony always aims at producing addressee-belief in the proposition testified to. One attraction of this view is that there clearly are many cases in which one person tells someone something in order to bring about a certain belief in them. For example, if I am running late to catch a train and ask someone on the street for the fastest way to get to the station, all I may care about is that I form an accurate belief as to the quickest route to the station. In such a case it just so happens that testimony is the most convenient way of achieving this end. If it happened to be more efficient for the stranger to hand me a map, presumably they would do so.

The simple view would be a non-starter were it not for the fact that the aim of testimony is not meant to include every part of testimony but rather what the parts of testimony are teleologically organized around. It is thus compatible with a wide range of theories about how testimony achieves its end, for instance that it does so by providing the addressee with a special, interpersonal reason or that it does so by serving as the basis of an inference to the truth of the proposition stated. The view is just that the point of testimony *as such* is getting them to believe what you say and such a view is compatible with a wide range of explanations of how such a testimonial belief is formed.

A useful example of a proponent of the simple doxastic view is Grice. Grice famously stressed the teleological character of communication by attributing to the speaker a complex self-reflexive intention. For Grice (1957), in uttering a proposition, a speaker makes manifest the intention to produce a belief in an addressee, a belief which is to be grounded in the addressee's recognition of the speaker's intention. Perhaps one reason for Grice's view's enduring influence on discussions of assertion, testimony, and conversation generally is that the simple doxastic view is very natural. There is something right about the idea that in telling you something I seek to make a difference in your mental life, a difference that it is reasonable to think must be related to your attitude towards the proposition I utter. On top of belief being the obvious candidate for this attitude, the simple doxastic view gains credibility from the fact that testimony is undeniably a good reason to form a belief. If I am asked why I believe something, a valid reason is: so-and-so told me.

Finally, there is a linguistic motivation for the simple doxastic view. The word "tell" seems to apply only to cases where the addressee does not know the truth of the proposition being uttered. If I believe *p* and you know that I believe *p* it is odd for you to nonetheless attempt to "tell" me *p*. "I know that!" is a valid complaint if one is told something one knows. But this seems to suggest that telling can aim at nothing more than belief. For if it did it would be reasonable for you to tell me something I believe: you would be aiming to achieve whatever it is that telling achieves over and above mere belief. But this does not make sense and so the doxastic view seems to follow.[114]

[114] This line of reasoning assumes that "tell" used colloquially tracks what philosophers mean by "testimony." I address this assumption later; for now the point is just to see the basic appeal of the simple doxastic view.

The view is also espoused more or less explicitly in much writing on testimony. For instance, the first condition of C. A. J. Coady's (1991) definition of testimony, in his seminal work *Testimony: a Philosophical Study*, is that the speaker's statement is offered as evidence that *p*. If evidence is understood in the usual way as simply that which justifies belief,[115] Coady seems to endorse the simple doxastic view: A statement in *p*, in order to be testimony, must be offered as (we could say: aim to be) the justification for a belief in *p.* In a similar vein, in influential work on testimony, Fricker (2006) has written that the characteristic import of a testimonial speech act is that the hearer can form a belief in what is said and that such acts are "consummated" when the hearer forms the relevant belief.

The arguments for the simple view, however, are not immune to scrutiny. While no one can deny that testimony is ordinarily a legitimate ground for belief, it does not directly follow from this that testimony as such aims at belief. The mere expressions of belief considered earlier are also, all things being equal, a reason for any one in earshot to believe what is said, but by hypothesis these do not aim to bring about any particular attitude. And the fact that it is infelicitous to use the word "tell" in cases where the addressee knows what is said is not decisive either as this may be an artifact of language with an explanation other than that the word "tell" tracks the philosophical concept of testimony and that testimony aims at belief.

Finally, one might wonder whether the simple doxastic view neglects the important role that the speaker plays in the aim of their own action. One might think that, for a speaker, testimony is not just about having the addressee come to bear some relation to what is uttered but about herself and the addressee coming to have a shared relation to that stated fact. This line of thinking suggests a more nuanced doxastic view on which the function of testimony is to bring about a state of affairs where speaker and addressee both believe p, believe that they believe p,

[115] See for example Kim (1998)

believe that they believe that they believe, and so on. I'll call this the *complex doxastic view*, that testimony in p aims at the parties to the testimonial exchange having common knowledge of p.[116] According to this view it is true that testimony aims at the addressee forming a single belief but only insofar as this belief is an aspect of the testified to fact being commonly believed by speaker and addressee.

The complex doxastic view runs parallel to work mentioned earlier on the aim of assertion. Testimony seems to be oriented towards a shared rapport between speaker and addressee. When I tell you something I am not just narrowly concerned with your beliefs nor with my own: I am seeking to make a fact out in the open, as it were. If one glosses a fact's being out in the open as its being common knowledge between two or more parties, as it often is, then our final worry with the simple doxastic view recommends the complex doxastic view.[117]

3. Objecting to Doxasticism

In this section I want to pose an objection to doxasticism as a whole, illustrating how it destabilizes both the simple and complex versions. While the objection is not decisive, I'll argue it is serious enough to warrant looking for an alternative view. In short, the problem is the existence of a certain sort of situation, one where it is rational for an agent to provide testimony to someone but whose rationality doxasticism does not have the conceptual resources to describe. Here is an imagined example:

> *John and Claire:* A couple, John and Claire, are at a couples therapy session. The couple is working through a betrayal: John has been lying to Claire about something for a long time and Claire is talking about the effect its discovery has had on her. She describes how she found out, the emotions it evoked, facts about her past that frame her perception of

[116] I am using the term common knowledge (as opposed to common belief) in the way David Lewis (1969) defined it, as the speaker and addressee believing what is said, believing they believe it and so on.
[117] Heal (1978) for instance defines publicity this way.

> the situation, and so on. As it happens, she addresses all this to the therapist; it happened to be the therapist's prompting that invited her to discuss these facts. John and Claire are sitting next to each other but Claire is facing the therapist who sits across from them. Something occurs to the therapist as Claire is speaking: that while Claire knows John is there, she seems to find it easier to explain herself as she faces the therapist. She seems to be avoiding John as she explains what really has more to do with him. The therapist stops her and says, "You are telling all this to me but perhaps you should tell it to John."

In this scene Claire, the therapist, and John all know that John is sitting beside Claire and know that he is paying attention to what she is saying. They know this and they all know that they know it and so on. Thus, in both the scenario described and the hypothetical one where Claire addresses John we can expect there to be common knowledge of the facts to which Claire testifies and, of course, first order beliefs in those facts as well. Yet, we have the intuition that Claire and John have something to gain by speaking to one another.

Before going into the details in the next section I want to provide an overview of how this case provides the basis for an argument against doxasticism. The key datum I want to read off from this scenario is the following: there are instances in which testimony from a speaker to an addressee in a proposition brings it about that the speaker and addressee stand in some sort of collective relationship to that proposition that cannot be understood *merely* in terms of the speaker and addressee having gained certain beliefs or formed common knowledge in that proposition. Now, doxasticism is a view about the aim of testimony and not what testimony brings about in any given case. So doxasticism is consistent with what we can straightforwardly read off from the case; however, deeper reflection on this relationship (between speaker, addressee, and the proposition) which testimony is capable of bringing about reveals that it is the kind of phenomenon that a view about testimony's aim ought to have an account of. The thought is that it is incumbent on a view like the doxasticist's to provide some sort of explanation of why it is rational for Claire to be prompted to address John specifically when John was already in a position to overhear her. I'll argue that, as it turns out, there just is no good explanation forthcoming from the doxasticist. The heart of my argument, then, consists in rejecting the most appealing of these attempts. While it is impossible to consider every possible one the doxasticist might essay, it seems that every natural avenue of explanation faces serious issues. My conclusion is that there is inference to the best explanation to be made, an inference to a non-doxastic view.

Explaining *John and Claire* comes out to answering the following question: holding all else equal, is learning a fact through being addressed by a given speaker different from learning it by overhearing that same speaker's testimony to another similarly situated hearer and if so how? The impression that certain theories of testimony may not be up to this task has been suggested before, but this case adds to that pressure by placing a further constraint on theories of testimony.[118] The difference between learning by being told and learning by overhearing cannot just be any difference but must come to something that could have a meaningful impact on human relationships like John and Claire's. Insofar as they merely allow themselves to be overheard by each other, the couple is not missing out on just anything but something crucial, something which matters to the well-being of their relationship. In other words, it may turn out that even theories of testimony expressly constructed to mark out addressees as standing in a special testimonial relationship to speakers fail to do justice to the details of this case.

The first order of business is to establish that *John and Claire* points to some genuine

[118] For examples of this sort of criticism see McMyler 2013 and Moran 2018.

phenomenon whose underlying structure we can inquire into. Perhaps the therapist's suggestion is just misguided, in which case there is no need for a philosophical account of the significance of address and no need to doubt the doxastic view. I take it that most advocates for doxasticism will not go this way and grant that the case has the ring of plausibility. For good measure, though, here are a few other examples of situations where an individual seems warranted in wanting to learn some fact specifically via a certain speaker's testimony:

> *Doctor-Patient:* A doctor decides that a patient will need to undergo a risky surgery and lets the patient in on it merely by telling a nurse in a loud enough voice that the patient will overhear. The patient seems to have a right to resent the doctor's mode of communication.
>
> *Dying Mother:* A dying mother wishes to communicate the contents of her will to her son directly as opposed to merely having her son discover its contents when she passes away.
>
> *Diary:* A husband asks to hear about his wife's day but is upset when she, busy with something else, merely hands him her diary to read instead. The husband feels the wife has missed something about what he was after in inviting her to speak about her day.

The need to explain the significance of address emerges so long as any one of these cases is believable. Indeed it is hard to deny that on occasion we care about confronting someone face to face and that sometimes these confrontations consist of nothing more than the exchange of information. Whether one can maintain this idea and doxasticism simultaneously remains to be seen but denying it outright is unjustified.

We can leverage the multiplicity of examples against another attempt to sweep the problem under the rug. In particular we can leverage it against the thought that Claire is simply being irrational. The doxasticist may be tempted to think, in other words, that Claire has an arbitrary, and therefore uninformative, fear of addressing people in her life about certain facts even when she is happy to have them come to believe these facts in another way. If that were true, the case's plausibility could be explained away as the therapist simply inviting Claire to tackle some sort of strange phobia.

The problem with this response is that this phobia or hang-up is something we can easily picture across a broad range of contexts. It is understandable, in other words, for a psychologically healthy person to have difficulty addressing another about a sensitive topic. Redescribing the case this way– in terms of a reasonable psychological difficulty– leaves in place the question the doxasticist has yet to answer: how is it that address puts speaker and addressee in a distinctive relationship, one that in certain situations can be stressful for a psychologically healthy adult?

Another attempt to evade the need for an explanation of the therapist's suggestion appeals to the thought that the therapeutic context itself makes address important in some way that is exceptional and irrelevant to philosophical analysis. But once more, the fact there are similar cases that take place in non-therapeutic contexts– such as *doctor-patient, diary, and dying mother*– blocks this deflationary response.

There is another response worth giving to address the worry about the setting of *John and Claire*: the therapeutic context is arguably a uniquely *good* environment in which to study our intuitions about testimony. The therapist's job is to get Claire and John to relate in the right way, to express themselves, and to cultivate the right kind of conversation between them. These notions (relation, expression, conversation) are intimately connected to the concept of testimony. Testimony is a basic way of relating to others; in testimony one expresses one's point of view; and one does so in the context of a conversation. Our intuitions about the interpersonal resources available to John and Claire, therefore, are closely connected with our intuitions about what speech as such can do.

The doxasticist may also try to resist the burden of explanation in the first place. We have established that the imagined scenario shows us something about what testimony, at least in some cases, can achieve or accomplish, but how are we to know that this accomplishment just is the aim of testimony? The view of the doxasticist is simply about this more limited topic: the state of affairs that testimony as such is directed at. It is consistent with any such view that testimony tends to have all sorts of knock-on effects antecedent or subsequent to its aim i.e. to the addressee believing what is said or to the addressee and speaker having common knowledge. Perhaps the doxasticist will suggest that that testimony is being exploited to achieve some effect alternative to its main aim just as a speaker might assert a random fact in order to give a hearer an indication of their voice.

Everything just said on behalf of the doxasticist can be conceded. What is crucial to our case is just that they feel a need to have a story about the significance of address. Let us consider our imagined doxasticist's own example: testimony, in many cases, gives an addressee an indication of the speaker's tone of voice. Even though this is obviously not the point or function of testimony it is the kind of thing the doxasticist should have no trouble accounting for. The doxasticist can simply say that the aimed at testimonial beliefs are grounded in the speaker's speech act and that this speech acts is often heard as opposed to being read or seen (in the case of a physical gesture) by the addressee. Were the doxasticist not able to provide such a story of how addressees frequently learn of a speaker's accent and some other view did we would have reason to doubt their view.

There is an abstract similarity between the function telling is performing in all of the cases we have seen so far: testimony as opposed to merely overhearing testimony puts speaker and addressee in some sort of direct encounter that, for one reason or another, is desirable to the parties involved. That there is a common thread running through these different cases suggests that whatever needs explaining in *John and Claire* is not an accident of the case but a characteristic effect of testimony– something achieved just in virtue of testimony being given and accepted. A view of testimony's aim ought to be able to account for such characteristic effects. While the doxasticist may or may not want to include this effect in the account of testimony's aim itself it seems that their account ought to make room for it. In what follows, I argue that there really is no such room.

There are a number of ways for the doxasticist to explain what is going on in cases where a speaker or addressee may feel that it is important for the speaker to address the addressee specifically. They can be sorted into two camps according to whether or not they conceive of the significance of testimonial address as something that is connected to the actual proposition testified to or not. I refer to the latter kind as "additive" explanations since they invoke some notion external to the addressee's relation to the testified to proposition, and I call the former "expansive" since they seek to expand how we think of an addressee's relation to the testified to proposition.

There are trivial differences between the beliefs one forms on the basis of address and on the basis of overhearing someone. If Claire tells John the things on her mind, he not only comes to believe her report in the good case, he believes he is being addressed by her, he believes his beliefs to be formed on the basis of her speech act addressed to him, he believes Claire is looking at him, and speaking in his direction. Unfortunately for the doxasticist, none of these beliefs seem like the kind of thing that could make an important difference in Claire and John's relationship. It does not seem that John's having any one of these additional beliefs is intrinsically desirable from the perspective of the health of their relationship.

A more promising thought echoes a worry we already dealt with, that a speaker communicates to their addressee some fact additional to the proposition testified to. Previously we considered and rejected the idea that this might be due to the therapeutic context but one might think that testimony as such contains some additional communicative content. Perhaps Claire's addressing John communicates something like, "I respect you," not as a matter of circumstance but just in virtue of being testimony. This added communicative dimension will have to be fairly minimal if it is to be present in every situation where testimony is given. Still, the reason this thought gives the doxasticist hope is that expressions of respect do seem like the kinds of things that might matter to an intimate relationship like John and Claire's.

I see two problems with this sort of approach. First, such an approach requires us to divide ordinary cases of telling into two parts in a way that is ad hoc and unintuitive. Surely it is more attractive to think that testimony in *p* just comes out to the communication of *p* (however we spell that out) as opposed to the communication of *p* plus some additional content. Secondly, and perhaps more importantly, we can easily imagine that Claire has been addressing John that day and will continue to do so after the therapy session. This means that we can imagine John already knows that Claire respects in this special testimonial way.

The problem for the doxasticist with appealing to some added communicative dimension is that intuitively the thing the couple is missing out on (when Claire merely allows John to overhear her) has to do with the very issues Claire is talking about: the betrayal, her feelings about it, and the effects it had on her. It is these facts which John and Claire need to get in the right orientation toward. What the different cases like *John and Claire* appear to have in common is not the communication of respect in the abstract but somehow respect in connection with the facts at hand.

Another additive explanation takes its cue from how we described the similarity between the different cases where address matters; there we suggested the idea of an encounter between addressee and speaker. In light of this, the doxasticist might suggest that in addition to bringing about beliefs in speaker and addressee, testimony provides the occasion for some kind of non-propositional encounter: a pure meeting of gazes perhaps or an irreducible moment of intersubjectivity. While I think this thought gets something right, these ways of describing such an encounter render it opaque and mysterious. The idea that through testimony there is a meeting of gazes is not an idea that the doxasticist can simply help themselves to. I'll suggest that it is only once we abandon the doxastic view that we are able to adequately describe the moment of intersubjectivity that this response is reaching for.

Considering the issues that additive explanations faced, the doxasticist may seek to make the significance of address more directly connected to the reason for belief that an addressee acquires through testimony. Indeed, the reader may be confused at this point as we have left until the end what is arguably the most straightforward thought about why Claire ought to address John, one which leaves doxasticism intact. The thought is that in addressing John about certain facts Claire signals to John that she is open to having a conversation with him about these very facts in addition to giving him a reason to believe what she says.

Surely the opening of a conversation has the weight of something of something with real therapeutic importance. Perhaps John and Claire have to have a dialogue about the betrayal that occurred, and perhaps Claire does not begin this dialogue if she only allows John to overhear her. Not only does this thought fit the details of the case, it also resonates with recent work on the concept of address. Benjamin McMyler (2013, 1064-1070) has argued that, for a given testimonial exchange, only the addressee is afforded a set of testimonial rights related to the speaker's testimony, in particular the right to defer questions from third parties to the speaker about the knowledge they gained from the testimony. Address partially determines what conversational moves are appropriate for an addressee both with respect to the speaker and other interlocutors and this power might explain why John really should be the addressee of Claire's speech.

The issue with this way of handling the case is that it does not fit with the other examples we gave. For example, for the doctor and the patient there may be nothing more important that needs to pass between them about the surgery the patient will undergo, and similarly for the mother on her deathbed, we can imagine that it is important to tell her son about her will even where there is no hope of them ever speaking again. Nor, on the face of it, does it seem important that John be able to respond to questions from third parties about Claire's testimony. There is something right in the idea that address makes conversational exchange possible, something which will have to be folded into any successful view of address, but the other examples we looked at suggest it is not a complete account of the target phenomenon.

Another theme in work which tries to understand testimonial address is that when a speaker testifies to an addressee that addressee and only that addressee gets a special reason for believing what is said, a reason which has been described as interpersonal or second-personal.[119] On these views John may form a justified belief on the basis of overhearing Claire talk to the therapist but it would have a different sort of ground than one based on her testimony to him. We can still think of testimony as aiming at belief, then, if we are open to the idea that the therapist

[119] See for example Moran 2018 and McMyler 2011.

thinks John ought to have a belief which is grounded in a special way, on a second personal reason or on an interpersonal reason.[120]

This response is less powerful than it first appears. It is just not clear why having a certain kind of reason, no matter how special, should be intrinsically valuable. The value of epistemic reasons seems to be exhausted by the beliefs they support: such reasons are worth something precisely in so far as they are reasons for beliefs. Something similar is arguably true for practical reasons: it may be the case that we have distinctly second personal reasons to act in certain ways, for example to fulfill promises and so on. However, it does not seem right to think that it is an end in itself to have such a reason. What is valuable are relationships, rights, and whatever generates practical reasons, not the reasons themselves. There's something artificial about John saying he would like to be Claire's addressee because he would like to have a special sort of reason. Thus, the account of the therapist's suggestion we are considering portrays John's request as fetishistic, even self-centered and this impression is borne out by this brief axiology of reasons.

4. Shared Knowledge as the Aim of Testimony

To arrive at a satisfying theory of testimony's aim we need to say more about what exactly we imagine Claire is avoiding when we imagine her shying away from speaking to John. One way of shedding some light on Claire's situation is to return to the motivation for the complex doxastic view. It was suggested that the simple view misses the sense in which testimony seeks to put some fact out in the open before the speaker and the addressee. We arrived at the complex doxastic view by glossing this openness as the fact being common knowledge, and we struggled to make sense of the couple's therapy case in these terms. But

[120] Moran (2018) for instance suggests that testimonial beliefs are based on the addressee's and speaker's mutual recognition of the speaker's intention.

perhaps the problem was not with the idea of openness itself but with our gloss of it. There's something right about saying that it is only when Claire speaks directly to John that the relevant facts will be out in the open between them and that this is what someone in Claire's position might want to avoid. Thus, if our objection to complex doxasticism is correct, the lesson of *John and Claire* is just that a fact being out in the open between people cannot be understood as that fact being common knowledge.

As it stands the phrase, "out in the open", while evocative, does not have a clear enough connection to traditional epistemological terminology to serve as the starting point for an epistemology of testimony. Where shall we go from here? One thing to notice is that the phrase "out in the open", like knowledge, is factive. It sounds wrong to say that a proposition is out in the open between agents if that proposition is false. Facts which are out in the open between agents are facts which those agents are aware of without necessarily taking any evaluative or affective stance towards them. It's natural to think, then, that there is some connection between the notion of knowledge and the notion of things being out in the open among agents. Indeed, another term we could have used to describe what Claire might want to avoid, acknowledgement, has a clear etymological connection to the term knowledge. Furthermore, notice that there is something Moore paradoxical about someone saying: *p* is out in the open between us but we do not believe it. Despite these striking connections between something's being out in the open and the concept of knowledge, we have seen, we cannot analyze it in terms of individual knowledge claims. I believe, therefore, that we can take a step towards clarity by describing agents for whom *p* is out in the open as having shared knowledge of *p*.

The language of sharing knowledge raises a new set of questions, but hopefully these are just the questions a theorist of testimony needs to consider. What is different about two agents sharing knowledge of *p* rather than simply separately knowing *p* and knowing that they know it and so on? There's two claims I wish to make in order to bring out this difference. The first is on the distinctive role shared knowledge plays in social life and the second is on the relationship between individual knowledge and shared knowledge.

5. Claim #1: Shared Knowledge Rationalizes Joint Action

It is widely accepted that there is a distinction to be made between acting alone and acting with others.[121] If, while walking down the street, I happen to find myself walking side by side with another person for a few blocks, this is different than the two of us walking down the street *together* even though the two may be indistinguishable for a period. In other words, just as a given action can be an exercise of individual agency so an action can be an exercise of shared agency. What is it in virtue of which we say an action is an exercise of collective agency?

One way of approaching this question is to ask: what is it about an action that makes it belong to one agent or another? What is it about my actions that makes me think of them as all belonging to a single person– as mine? Suppose I make breakfast, then do some work, then run an errand. Why do these count as a set of actions that have a common agential source? Once we have an answer to this question then we should be able to say why it is that some actions are properly attributable to groups or collectives.

Here's a natural answer. An action is attributable to a specific agent in virtue of it being that agent that is subject to a set of rational norms related to the action in question, norms of coherence, practical reasoning, and so on. An action belongs to: the person whose beliefs must cohere with that action, the person whose other actions must be compatible with the action in question, the person who is responsible for carrying out that action in a rational way.

[121] See, for instance, discussions of joint action by Schmid (2023), Laurence (2011), and Gilbert (1990).

We can call attention to these norms by, for instance, asking the agent "what are you doing?", "why are you doing that?", "how are you going to do that?" and so on. The agent of a particular action is the person to whom we would address these questions if we were to ask them or, simply, the person that is implicitly sensitive to them in carrying out the action in the first place. If all this is right then the question, in virtue of what does an action count as an exercise of collective agency, is really, what makes it the case that a collective is subject to a set of rational norms.

Consider what "being subject" to these rational norms demands in the individual case. It is not just that the individual is affected in one way or another by such norms but that they are aware of them in a way that has a direct bearing on their action.[122] If a hypnotist has me perform some action then, even though "what I'm doing" is in some sense "sensitive" to some rational norms– since the hypnotist may be sensitive to them– the action is not really attributable to me: it's not an exercise of my agency. Thus, an individual in order to act as an agent must be aware of what they are doing.[123]

What does all this mean for the collective case? It suggests that insofar as there genuinely are instances of *shared* action, it must be possible for multiple agents to have practical *knowledge* that is shared.[124] So agents who engage in joint activity need to have shared practical knowledge of what they are doing.[125] So far we have just rehearsed an argument for the presence

[122] I take it that Bilgrami (2006, 93-120) makes essentially this point.
[123] This point is argued for in various places, for instance in Small 2012, Marcus 2018, and Setiya 2016.
[124] For the same point see, Roessler 2020.
[125] A natural worry, at this point, is that this is an overly strong demand. There would seem to be plenty of actions done by groups (the army, corporations, etc.) where it would surely be stretching things to say that there is collective awareness– whatever that turns out to be. There obviously are cases in which individuals make up some sort of corporate agent (something which does things) but in which the individual members that make it up may not be aware of what the corporate agent as a whole is doing. I think there is a sense in which these cases are cases of group action; however, they are not cases of shared action in the sense we are interested in here. Group actions of corporate agents are not products of the collective agency of all involved. This is a welcome consequence since it clearly is the case that whomever it is that is at the top of the corporate hierarchy is more responsible for what a corporate agent does then the people that are merely carrying out their assigned roles. The leader or board of trustees etc. is responsible for it being the case that the corporate agent's actions fit within a single coherent world view and

of “shared practical knowledge”, that is findable in other work on joint agency.[126] To make such considerations relevant to our discussion of the kind of state generated by testimony we need to go one step farther. Testimony typically concerns matters of fact while practical knowledge is precisely not the kind of thing one agent can simply inform another of. Afterall, if you are in need of telling that we have shared practical knowledge then we do not have shared practical knowledge to begin with. Thus, we need to show that agents who act together also need something like shared empirical knowledge.

Why think this? First, consider what action is. To act is to bring about some change in the world. When I make coffee I alter the condition of my immediate environment to bring it in accordance with my wish for coffee. Such an alteration betrays my understanding of or my awareness of the state of my environment prior to action: I know where the kettle is, the beans are, etc. At the foundation of action, then, is empirical knowledge of one’s environment.

What about joint action? Granted, agents that act together each have their parts to do. If we’re making coffee together we might agree that I grind the beans and you pour the water. My grinding of the beans will reflect, among other things, my empirical knowledge of the beans while you're pouring the water will reflect your empirical knowledge of the location of the water and so on. And it does not seem to be the case that these respective bits of knowledge need to be anything more than individually held.

However, consider what makes such coordinative behavior possible in the first place: we had to decide together what each of us was going to do. This step was crucial to our being entitled to go off and perform our parts. Joint action involves multiple agents, and, even in such a

harmonizes with other of its actions. Moreover, this difference between leader and subordinate is one that has a pre-theoretic legitimacy. In legal contexts we recognize a difference between being entangled with a criminal enterprise and being its mastermind.

[126] Again see Roessler 2020 or Rödl 2018.

situation where joint action involves more or less identical parts, it is a fact of nature that no two agents can do *exactly* the same thing. Thus, all joint action requires the parties to decide implicitly or explicitly, how it is that they will go about doing what they have decided to do.[127] This question *always* arises for agents engaged in joint action. Even if the decision we must make is as simple as: I do my part *over here* and you do yours *over there*, still there is some decision that has to be made about how it is that we go about acting jointly. Thus, joint action always involves what we might call *interagential specification*.

Here is my claim: the empirical knowledge necessary for interagential specification needs to be shared. It needs to be shared rather than individually held for exactly the same reason that practical knowledge must be shared. Let us return to what makes a particular agent the agent of a given action. Recall, we said that the agent of a given action is the agent to whom we would address certain questions about the rationality of their action. Imagine we have decided to make coffee: you grind and arrange the beans, I prepare and pour the water. Suppose this is our agreement and you set about placing the beans in the dishwasher. An onlooker would be puzzled. To whom should they address their puzzlement? You. You are the one placing the beans in the dishwasher. Preparing the beans is your part and the empirical knowledge needed to do so must be had by you. So, here, as expected, there is a correspondence between the agent who ought to have some bit of empirical knowledge and the agent whose action depends on such empirical knowledge. But suppose, instead, that you and I decide we will make coffee by your grinding and preparing the beans while I grind and prepare another set of beans. An onlooker would be puzzled, but their puzzlement would only be appropriately addressed *to us*. Why are you two making coffee like that, they might reasonably ask.

[127] Bratman (2014) makes the same point, when talking about meshing sub-plans.

The fact that agents of joint action are the collective addressee of certain questions regarding the empirical knowledge undergirding action implies that they have shared empirical knowledge. The fact that interagential specification is a fundamental feature of joint action shows that shared empirical knowledge is a fundamental feature of joint action. Hence, one thing to say about shared knowledge is that it is the kind of state that rationalizes joint activity.

6. Claim # 2: The Transparency of Shared Knowledge

The second idea I want to put forward about shared knowledge has to do with the cognitive relation each agent bears towards the content of shared knowledge they are a party to. To do so, I want to draw on some recent work on first personal self-consciousness, in particular work that addresses the transparency of first personal thought. An influential observation about self-knowledge is that I can know whether I believe *p* by considering *p* itself.[128] How is it that considering *p* puts me in a position to know about an ostensibly separate topic, namely my own mind? A lot has been said in response to this question. What will be useful for our purposes is a distinction Matt Boyle (2024) brings to bear on it, namely between a thought's content and its mode of presentation.[129]

Suppose you see a cat lying on your lap and think "this cat is lying on my lap". Boyle argues that the significance the perceptual demonstrative, "this", has in such a thought cannot be understood in terms of the content of the thought expressed. Thinking "this cat is lying on my lap" is not the same as thinking "the cat I currently perceive is lying on my lap", Boyle suggests. Rather the perceptual demonstrative marks a certain way in which the cat on my lap is present to me in thought; it marks the perceptual mode of presentation. As Boyle puts it "a feature of representation [belongs] to its *mode* of presenting its object… when this feature expresses some

[128] This point is most associated with Evans (1982).
[129] I am not suggesting that this distinction is Boyle's invention. My sense is that Boyle is indebted to Evans (1982) here.

aspect of the subject's cognitive relation to the represented object" (2025, 75).

Boyle suggests that while a representation's mode is not an aspect of its content, it is something the agent can make explicit through reflection. If I think of the cat in my lap as "this cat" my thought belies an implicit understanding of the fact that I am perceiving it, one that can be made explicit in further thought. Boyle suggests we can say something analogous about belief in order to explain transparency. When I believe *p* the content of my thought is *p*; however, I represent *p* in a distinctive way, i.e. "as believed by me". One can see that I am implicitly sensitive to this mode of presentation in virtue of the fact that I am sensitive to various rational pressures associated with *p*. Believing *p* commits *me* to believing anything entailed by *p* and refraining from believing anything inconsistent with it, for instance. Thus, I can answer questions about whether I believe *p* to be true by considering *p* because once I have settled whether *p* I am in a position to reflect on what is implicit in settling p: my belief or disbelief in *p*.

I want to stress the dialectical significance of this point to discussions of self-knowledge since I later will propose an analogous dialectical move concerning testimony. The larger aim of a project on self-knowledge like Boyle's is to give an account of what it means to be aware of one's mind "first-personally" rather than through observation. Pointing to the special sort of awareness one has of a representation's mode of presentation helps to satisfy this aim: it gives us a way of saying what is distinctive about first-personal self-knowledge.

Of course, there is much more that could be said here about self-knowledge. For now, what is relevant is the following idea: the consciousness an agent has of the mode of presentation of a given representation can contain the grounds for a further reflective, "agent-directed" thought. Let's see how this might bear on our topic. Instead of perceiving a cat purring, imagine jointly attending to a cat purring with your friend. Your attention, in such an episode, will be on the cat not on your friend. Moreover, we can say the cat will be presented to you in a distinctive mode, one which you could express to your friend by saying “this cat is purring”. Notice: this “this-thought” will have different truth conditions from the one in the individual case. Here, a “this-thought”, and sentences with “this” in them, will be appropriate only insofar as your friend also perceives the cat. This suggests that your awareness of the cat in the context of joint attention, while not making explicit references to your friend, is still an awareness of the cat which presupposes that your friend also perceives it.[130]

A similar analogy holds between individual belief and the content of a conversation between two parties. We suggested earlier that my belief *that the cat is purring* does not involve an explicit awareness of myself as believer. Rather, it involves an implicit awareness of its rational relationship to other rational attitudes I hold, and in this way presupposes first personal self-awareness. A similar point holds for conversation. Imagine you and I are talking about how cute the purring cat is. Here, our focus is on the cuteness of the purring cat. And yet, insofar as this is a conversation between us, we will each feel the force of certain rational relationships between the content of our conversation and the other attitudes we hold. If you had said previously that no cat was cute, I would now have grounds for pointing out a tension between what you believe and what we have agreed on. This suggests that in a conversation facts are presented to me in a certain mode, one which presupposes that me and the other participants are committed to their truth. Hence, there is a kind of transparency in conversation. Conversations have an explicit topic but they put each party to it in a position to reflect and make explicit awareness of one another.

If the aim of testimony is shared knowledge then we can draw on this point about the transparency of conversation to illuminate shared knowledge. In particular, we can draw on the

[130] For a view of joint attention along these lines see Campbell 2005.

distinction between content and mode of presentation to shed light on the relationship each agent bears towards one another when they have shared knowledge. When agents share knowledge of *p* the content of their thought is simply *p*, but it is present to them in a distinct way, namely as believed by us. Moreover, each party to shared knowledge is in a position to reflect on their awareness of *p* to make explicit their understanding that: "we know *p*" as well as "I know *p*" and "you know *p*". Moreover, when they have this kind of awareness of one another, their awareness is not impersonal or observational. Just as one knows one's own mind in a special first personal way so agents to shared knowledge know one another's mind in a special interpersonal way.

7. Interpersonal Priorism: An Epistemological Account of Testimony

Let's return to testimony. We said that epistemologists tend to ask: what sort of reason for belief does the audience of testimony receive? If the aim of testimony is shared knowledge we need to ask a different question: how is it that a testimonial act can justify the generation of shared knowledge? Each of these questions directs our attention to a different set of knowledge claims. The first framing directs our attention to the epistemic situation of the hearer of testimony: how can a speaker make a difference to what the hearer knows? The second framing, our framing, is to ask what is it about speaker and hearer together which legitimates their shared knowledge in the wake of a testimonial act? A natural thought is that there must be some shared knowledge– not itself grounded in testimony– that obtains prior to the testimonial act, which itself serves as the basis of its successful execution.

To see how such a view might work, consider the wider context in which testimony is given. One tells another some fact almost always in the context of a larger conversation, a context which includes a variety of shared expectations, meanings, and presuppositions. Indeed, speakers always preface their testimony in some way or another. One almost never simply begins asserting a fact to another but instead always introduces oneself as a speaker in some way or another. This introductory remark is often a simple greeting. I want to suggest that such prefatory remarks are more than superficial niceties. They play a foundational role in testimonial exchanges by establishing a space of shared knowledge. The content of this initial shared knowledge, I suggest, includes basic facts like: "you are there", "I am here", "we are speaking English", "I see you", "you see me", and so on. In other words, it initially includes facts which would not require either party to attest to them.

The significance of this initial shared space to the debates around interpersonalism comes out of our earlier discussion of transparency. We said that each party to shared knowledge is in a position to make explicit to themselves that the other knows what is shared knowledge. As we saw with our discussion of self-knowledge, such other-directed awareness constitutes a special more immediate awareness of the other than the kind of awareness an onlooker or overhearer might have. It is an awareness that is the same kind of awareness that characterizes one's knowledge of one's own mind. This is the key point: when agents have shared knowledge, each has a direct acquaintance with the other party's epistemic capacities. It is not just that each party knows, observationally, the other is a competent knower. Each knows of the other's competence in a special interpersonal way. In this way, prior to any testimonial act, each party has already recognized the other as a competent epistemic agent and therefore a legitimate offerer of testimony. Moreover, they do so not just as an impersonal judge of epistemic agents– as an overhearer might– but in a distinctly interpersonal way, namely through reflection on their own relationship to the world.

I call the above view "interpersonal priorism". Let me say why I think interpersonal priorism constitutes an improvement over the alternatives in the literature. On the one hand, interpersonal priorism fully vindicates the intuitions behind interpersonalism. Testimonial acts succeed specifically on the basis of an interpersonal relationship. While overhearers can ascertain the reliability of a speaker and believe the contents of the speaker's testimony as a consequence, they would do so on a fundamentally different basis than the intended addressee. In contrast to the overhearer, the addressee's appreciation of the speaker's capacities is an aspect of the addressee's own mode of representing the world rather than merely an aspect of its content. However, this interpersonal view fully answers Lackey's objection. Background considerations of reliability and competence– via recognition of a speaker's legitimate status as epistemic agent– (a) play a central role in our story and (b) do so in such a way that obviates the need to posit a special, additional interpersonal reason that matters epistemically over and above such considerations. It is the manner in which considerations of reliability, sincerity and so on are grasped by a hearer that gives knowledge from testimony its distinctive interpersonal flavor rather than any special, self-standing reason. Finally, interpersonal priorism makes perspicuous the non-evidentialist position as we defined it earlier. Testimony is not evidence since testimony's primary aim is not the generation of individual knowledge but shared knowledge.

Before concluding let us consider a natural objection to our view. One might worry that our view fails to make sense of cases where the speaker's goal really is the simple transferring of information. Consider the simple act of asking a stranger for the time and receiving it. I accept, of course, that there are such cases. What I want to say is simply that, in such cases, speaker and hearer enter into a conversation simply for the sake of exploiting the fact that it affords the possibility of one party transmitting knowledge to the other via shared knowledge. It would be a problem for our view if such cases were paradigmatic. But I do not think that they are. For the most part, I submit, ordinary testimonial exchanges do have the function of coordinating the speaker and hearer rather than achieving a simple transference of information.

8. Conclusion

I have argued that the aim of testimony is shared knowledge, a state which rationalizes joint action and which puts each party in a state to know the other's mind through reflection. Through such knowledge, parties to a conversation are directly acquainted with the epistemic capacities of one another. The view that testimonial belief is belief grounded in such acquaintance vindicates the interpersonalist program. On this picture testimony is not evidence since it does not typically aim at supporting individual belief. Furthermore, testimonial knowledge has a distinctly interpersonal source not because it is based on a special interpersonal reason but because it is based on ordinary facts– such as the speaker's credibility and reliability– that are grasped in a distinctly interpersonal way.

References

Akkoyunlu, Eralp A., Ekanadham, Kattamuri, and Huber, R.V.. 1975. “Some Constraints and tradeoffs in the Design of Network Communications.” *SIGOPS Operating Systems Review,* 9: 67-74.

Alonso, Facundo M. 2009. “Shared Intention, Reliance, and Interpersonal Obligation.” *Ethics* 119:444-75.

———. “What is reliance?” *Canadian Journal of Philosophy* 44(2):163-183.

———. 2024. “The Self-Reinforcing Nature of Joint Action.” *Philosophical Studies*, 181:1047-1065.

Anscombe, G.E.M. 1979. “What Is it to Believe Someone?”. In C.F. Delaney, ed., *Rationality and Religious Belief*: 141-151. Notre Dame: Notre Dame University Press.

Arpaly, Nomy and Schroeder, Timothy. 2014. *In Praise of Desire*. Oxford:Oxford University Press.

Aumann, Robert. 1987. “Correlated Equilibrium as an Expression of Bayesian Rationality.” *Econometrica*, 55: 1-18.

Austin, John L, 1962. *How to Do Things With Words*. Cambridge, MA: Harvard University Press.

Bacharach, Michael. 1999. “Interactive team reasoning: A contribution to the theory of co-operation.” *Research in Economics* 53: 117-147.

———. 2006. *Beyond Individual Choice*. Natalie Gold and Robert Sugden, eds., Princeton: Princeton University Press.

Baker, Judith and Philip Clark. 2018. “Epistemic Buck-Passing and the Interpersonal View of Testimony.”*Canadian Journal of Philosophy* 48 (2): 178-199.

Baier, Annette. 1986. “Trust and Antitrust.” *Ethics* 96(2):231-260.

Benton, Matthew A. 2025. “The Epistemology of Interpersonal Relations.” *Noûs* 59 (1):92-111.

Bilgrami, Akeel. 2006. *Self-Knowledge and Resentment*. Cambridge: Harvard University Press.

Binmore, Ken and Samuelson, Larry. 2001. “Coordinated Action in the Electronic Mail Game.” *Games and Economic Behavior* 35:6-30.

———. 2008. “Do Conventions Need to Be Common Knowledge?” *Topoi* 27:17-27.

Boyle, Matthew. 2012. “Essentially Rational Animals.” In Günter Abel and James Conant, eds., *Rethinking Epistemology*: 395-428. Berlin: Gruyter.

———. 2024. *Transparency and Reflection: A Study of Self-Knowledge and the Nature of Mind*. Oxford: Oxford University Press.

Bratman, Michael. 1987. *Intention, Plans, and Practical Reason*. Cambridge: Harvard University Press.

———. 2014. *Shared Agency: A Planning Theory of Acting Together*. New York: Oxford University Press.

Brennan, Johnny. 2021. “Recognition trust.” *Philosophical Studies* 178(11):3799-3813.

Campbell, John. 2005. “Joint attention and common knowledge” in Naomi Eilan, Christoph Hoerl, Teresa McCormack & Johannes Roessler (eds.) *Joint Attention: Communication and Other Minds: Issues in Philosophy and Psychology*. Oxford: Clarendon Press.

———. 2018. “Joint Attention.” In Marija Jankovic and Kirk Ludwig, eds., *The Routledge Handbook of Collective Intentionality*: 115-129. New York: Routledge.

Carter, Adam J. 2024. “Therapeutic Trust.” *Philosophical Psychology* 37(1): 38-61.

Chwe, Michael. 2001. *Rational Ritual: Culture, Coordination, and Common Knowledge*. Princeton, NJ: Princeton University Press.

Clark, Herbert and Marshall, Catherine. 1981. “Definite reference and mutual knowledge.” In A. Joshi, B. Webber, and I. Sag, eds., *Elements of Discourse Understanding*. Cambridge: Cambridge University Press.

Coady, C.A.J. 1992. *Testimony: A Philosophical Study*. Oxford: Oxford University Press.

Colman, Andrew M. and Gold, Natalie. 2017. “Team reasoning: Solving the puzzle of coordination.” *Psychonomic Bulletin & Review* 25:1170-1783.

———. 2020. “Team Reasoning and the Rational Choice of Payoff-Dominant Outcomes in Games.” *Topoi* 39:305-316.

Colombo, Camilla and Guala, Francesco. 2023. “Rational Coordination Without Beliefs.” *Erkenntnis* 88:3163-3178.

Cogley, Zac. 2012. “Trust and the Trickster Problem.” *Analytic Philosophy*, 53(1):30-47.

Craig, Edward. 1990. *Knowledge and the State of Nature*. Oxford: Oxford University Press.

Cubitt, Robin, and Sugden, Robert. 2003. “Common Knowledge, Salience, and Convention: A Reconstruction of David Lewis’s Game Theory.” *Economics and Philosophy* 19: 175-210.

Dancy, Jonathan. 2000. *Practical Reality*. Oxford: Oxford University Press.

Darwall, Stephen. 2017. “Trust as a Second-Personal Attitude (of the Heart).” In Faulkner and Simpson: 35-51.

D’Cruz, Jason. 2020. “Trust and Distrust”. In Simon: 41-51.

de Kenessey, Brian. 2020. “Promises as Proposals in Joint Practical Deliberation.” *Noûs* 54(1): 204-232.

———. 2022. “Promises, Offers, Requests, Agreements.” *Ergo an Open Access Journal of Philosophy* 9: 34.

Domenicucci, Jacopo and Holton, Richard. 2017. “Trust as a Two-Place Relation.” In Faulkner and Simpson: 149-161.

Doody, Ryan. 2019. “If There Are No Diachronic Norms of Rationality, Why Does it Seem Like There Are?” *Res Philosophica* 96(2): 141-173.

Ebels-Duggan, Kyla. 2008. “Against Beneficence: A Normative Account of Love.” *Ethics* 119(1):142-170.

Evans, Gareth. 1982. *The Varieties of Reference*. Oxford: Oxford University Press.

Fagin, Ronald, Halpern, Josephy Y., Moses, Yorman & Vardi, Moshe Y. 1995. *Reasoning about Knowledge*. Cambridge, MA: MIT Press.

Faulkner, Paul and Simpson, Thomas, eds. 2017. *The Philosophy of Trust*. Oxford: Oxford University Press.

Faulkner, Paul. 2011. *Knowledge on Trust*. Oxford: Oxford University Press.

———. 2015. “The Attitude of Trust is Basic.” *Analysis* 75(3): 424-429.

———. 2017. “The Problem of Trust.” In Faulkner and Simpson:110-129.

———. 2025. *Trust and Cooperation*. Oxford University Press.

Ferrero, Luca. 2022. “Structures of Temporally Extended Agency.” In Carla Bagnoli, ed., *Time in Action: The Temporal Structure of Rational Agency and Practical Thought*. New York: Routledge.

Frankfurt, Harry. 1971. “Freedom of the will and the concept of a person.” *Journal of Philosophy* 68(1): 5-20.

Fricker, Elizabeth. 1987. “The Epistemology of Testimony.” *Aristotelian Society Supplementary Volume* 61: 57-83.

———. 2006. “Testimony and Epistemic Autonomy.” In Jennifer Lackey & Ernest Sosa, eds., *The Epistemology of Testimony*: 225-250. New York: Oxford University Press.

———. 2019. “‘Believing the Speaker’ Versus Believing on Evidence: A Critique of Moran.” *European Journal of Philosophy* 27(3): 767-776.

Gauthier, David. 1975. “Coordination.” *Dialogue* 14: 195-221.

Gauthier, David. 1994. “Assure and Threaten,” *Ethics* 104: 690-721.

Gilbert, Margaret. 1989. *On Social Facts*. New York: Routledge.

———. 1990. “Walking Together: A Paradigmatic Social Phenomenon.” *Midwest Studies in Philosophy* 15(1):1-14

———. 1993. “Is an Agreement an Exchange of Promises?” *The Journal of Philosophy* 90: 627-649.

Godard, Emmanuel and Perdereau, Eloi. 2021. “Back to the Coordinated Attack Problem.” In Pierre-Louis Curien, ed., *Mathematical Structures in Computer Science*: 1089-1113. Cambridge: Cambridge University Press.

Gold, Natalie and Sugden, Robert. 2007. “Collective Intentions and Team Agency.” *The Journal of Philosophy* 104:109-137.

Goldberg, Sanford C. 2020. “Trust and Reliance. ” In Simon: 97-108.

Gray, Jim. 1978. “Notes on Data Base Operating Systems.” In R. Bayer, R.M. Graham, G. Seegmüller, eds., *Operating Systems, An Advanced Course*: 393-481. New York: Springer-Verlag.

Graham, Peter J. 2023. “Proper Functionalism and the Organizational Theory of Functions.” Luis R. G. Oliviera, ed., *Externalism about Knowledge*: 249-276. Oxford: Oxford University Press.

Greco, Daniel. 2023. *Idealization in Epistemology*. Oxford: Oxford University Press.

Grice, H.P. 1957. “Meaning.” *The Philosophical Review* 66(3): 377-388.

Grzankowski, Alex. 2016. “Limits of Propositionalism.” *Inquiry* 59(7-8):819-838.

Halpern, Joseph Y. and Moses, Yoram. 1990. “Knowledge and Common Knowledge in a Distributed Environment.” *Journal of the ACM (JACM)*, 37:549-587.

Harris, Daniel. 2020. “We Talk to People, Not Contexts.” *Philosophical Studies 177*:2713-2733.

Hawley, Katherine. 2014. “Trust, Distrust and Commitment”, *Noûs* 48(1): 1-20.

———. 2017. “Trustworthy Groups and Organizations.” In Faulkner and Simpson: 52-70.

———. 2019. *How to Be Trustworthy*. Oxford: Oxford University Press.

Heal, Jane. 1978. “Common Knowledge.” *The Philosophical Quarterly* 28:116-131.

Heape, Alexander. 2020. “Promises, obligation, and reliance.” *Philosophy and Phenomenological Research* 104(1):150-170.

Heeney, Matthew. 2019. “Diachronic Agency and Practical Entitlement.” *European Journal of Philosophy* 28:177-198.

Hieronymi, Pamela. 2008. “The Reasons of Trust.” *Australasian Journal of Philosophy*, 86(2): 213-236.

Hinchman, Edward S. 2005. “Telling as Inviting to Trust.” *Philosophy and Phenomenological Research* 70(3): 562-587.

Holton, Richard. 1994. “Deciding to Trust, Coming to Believe.” *Australasian Journal of Philosophy* 72(1): 63-76.

Hume, David. 1978. *A Treatise of Human Nature.* L.A. Selb-Bigge and P.H. Nidditch, eds. Oxford: Oxford University Press.

Jones, Karen. 1996. “Trust as an Affective Attitude”, *Ethics* 107(1): 4-25.

———. 2012. “Trustworthiness.” *Ethics* 123(1):61-85.

———. 2019. “Trust, Distrust, and Affective Looping.” *Philosophical Studies* 176(4): 955-968.

Julius, A.J. 2013. “The Possibility of Exchange.” *Philosophy and Economics* 12: 361-374.

Kelp, Christoph and Simion, Mona. 2023. “What is trustworthiness?” *Noûs* 57(3):667-683.

Kelsall, Joshua. 2024. “Towards a Non-Reliance Commitment Account of Trust.” *Journal of Value Inquiry*: 1-17.

Keren, Arnon. 2014. “Trust and Belief: A Preemptive Reasons Account.” *Synthese* 191(12): 2593-2615.

Korsgaard, Christine M. 2008. *Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford: Oxford University Press.

———. 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

———. 2018. *Fellow Creatures: Our Obligations to the Other Animals*. Oxford: Oxford University Press.

Kukla, Quill R. 2022. “Knowing things and going places.” *European Journal of Philosophy*, 31(1):266-282.

———. forthcoming. “Uptake and Refusal.” *Inquiry: An Interdisciplinary Journal Of Philosophy*.

Lackey, Jennifer. 2008. *Learning From Words: Testimony as a Source of Knowledge*. Oxford: Oxford University Press.

———. 2014. “A Deflationary Account of Group Testimony.” In Jennifer Lackey, ed., *Essays in Collective Epistemology*: 64-94. Oxford: Oxford University Press.

Lahno, Bernd. 2020. “Trust and Emotion.” In Simon: 147-159.

Lamport, Leslie, Shostak, Robert & Pease, Marshall. 1982. “The Byzantine Generals Problem.” In *ACM Transactions on Programming Languages and Systems* 4:382-401.

Laurence, Ben. 2011. “An Anscombian Approach to Collective Action.” In Ford, Hornsby, Stoutland, eds., *Essays on Anscomb's Intention*: 270-294. Cambridge, MA: Harvard University Press.

Lederman, Harvey. 2017. “Two Paradoxes of Common Knowledge.” *Noûs* 52(4):921-945.

———. 2018. “Uncommon Knowledge.” *Mind* 127:1069-1105.

Leonard, Nick. 2016. “Testimony, Evidence and Interpersonal Reasons.” *Philosophical Studies* 173:23333-2353.

———. 2018. “The Transmission View of Testimony and the Problem of Conflicting Justification.” *American Philosophical Quarterly* 55(1): 27-35.

Lepore, Ernie and Stone, Matthew. 2015. *Imagination and Convention: Distinguishing Grammar and Inference in Language*. Oxford University Press.

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

Loets, Annina J. 2021. "Qua Qualification." *Philosophers' Imprint* 21(27).

Ludwig, Kirk. 2020. "What is Minimally Cooperative Behavior?" In Anika Fiebich, ed., *Minimal Cooperation and Shared Agency*: 9-40. Cham: Springer.

Malmgren, Anna-Sara. 2006, "Is There A Priori Knowledge by Testimony?" *The Philosophical Review* 115(2): 199-241.

Marcus, Eric. 2018. "Practical Knowledge of a Normative Judgement." *Manuscrito* (4): 319-347.

Marušić, Berislav. 2015. *Evidence and Agency*. Oxford: Oxford University Press.

———. 2017. "Trust, Reliance, and the Participant Stance." *Philosophers' Imprint* 17:1-10.

———. 2024. "Interpersonal Reasoning: A Philosophical Psychology of Testimonial Trust." *European Journal of Philosophy* 33(2):531-549.

McDonald, Lucy. 2022. "Reimagining Illocutionary Force." *Philosophical Quarterly* 72: 918-939.

McGeer, Victoria. 2008. "Trust, hope, and empowerment." *Australasian Journal of Philosophy*, 86(2), 237-254.

McGeer, Victoria and Pettit, Phillip. 2017. "The Empowering Theory of Trust." In Faulkner and Simpson: 15-35.

McLeod, Carolyn. 2020a. "Trust." *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta & Uri Nodelman, eds., Fall 2023 Edition.

———. 2020b. "Damage to Trust." In *Conscience in Reproductive Health Care: Prioritizing Patient Interests*: 65-87. Oxford University Press.

McMyler, Benjamin. 2011. *Testimony, Trust, and Authority*. Oxford: Oxford University Press.

Merricks, Trenton. 2015. *Propositions*. Oxford: Oxford University Press.

Mintoff, Joe. 2004. "Is An Agreement An Exchange of Intentions?" *Pacific Philosophical Quarterly* 85:44-67.

Moran, Richard. 2005. "Getting Told and Being Believed." *Philosopher's Imprint* 5(5): 1-29.

———. 2018. *The Exchange of Words: Speech, Testimony, and Intersubjectivity*. Oxford: Oxford University Press.

———. 2019. "The Exchange of Words: Replies to Critics." *European Journal of Philosophy* 27(3): 786-795.

Morin-Martel, Alexis. forthcoming. "Full-Blooded trust as a Respectful Attitude." *Ergo*.

Morris, Stephen, and Shin, Hyun Song. 1997. "Approximate Common Knowledge and Co-ordination: Recent Lessons from Game Theory." *Journal of Logic, Language, and Information* 6:171-190.

Nguyen, C. Thi. 2022. "Trust as an unquestioning attitude." In Tamar Szabó Gendler, John Hawthorne, and Julianne Chung, eds., *Oxford Studies in Epistemology*: 214-244. Oxford:Oxford University Press.

Nozick, Robert. 1989. "Love's Bond". *The Examined Life: Philosophical Meditations*: 68-86. New York: Simon & Schuster.

Nickel, Philip J. 2007. "Trust and Obligation-Ascription." *Ethical Theory and Moral Practice* 10(3):309-319.

Paternotte, Cedric. 2011. "Being Realistic About Common Knowledge: a Lewisian Approach," *Synthese* 183:249-276.

———. 2017. "The Fragility of Common Knowledge." *Erkenntnis* 82(3):451-472.

Patrizio, R.F. 2023. "Trust's Meno Problem: Can the doxastic view account for the value of trust." *Philosophical Psychology* 37(1): 18-37.

Pease, Marshall, Shostak, Robert., & Lamport, Leslie. 1980. "Reaching Agreement in the Presence of Faults." *Journal of the Association for Computing* 27:228-234.

Putnam, Hillary. 1960. "Minds and Machines." In Hook, ed. *Dimensions of Mind*:138-164. New York: New York University Press.

Rachar, Matthew. 2023. "Conditional Intentions and Shared Agency." *Noûs* 58(1):271-288.

Reid, Thomas. 2002. *Essays on the Intellectual Powers of Man,* ed. Derek R. Brookes. University Park, PA: Pennsylvania State University Press.

Rimoldi, Florencia and Federico Penelas. 2025. "Testimony and Non-Evidential Reasons for Belief (A Non-Purist Place for Interpersonalism)." *Episteme*, 22, 283-303.

Roessler, Johannes. 2020. "Plural Practical Knowledge." *Inquiry* 67(4): 1067-1086.

Rödl, Sebastian. 2018. "Joint Action and Plural Self-Consciousness." *Journal of Social Philosophy* 49(1):124-136.

Ross, Angus. 1986. "Why Do We Believe What We Are Told?" *Ratio* 28(1): 69-88.

Roth, Abraham. 2014. "Prediction, Authority, and Entitlement in Shared Activity." *Noûs* 48(4):626-652.

Rubinstein, Ariel. 1989. "The Electronic Mail Game: Strategic Behavior Under Almost Common Knowledge." *The American Economic Review* 79: 385-391.

Salome-Sehr, Jules. 2024. "How to be a minimalist about shared agency." *Philosophy and Phenomenological Research* 109:155-178.

Sbardolini, Giorgio. 2024. "Who's Afraid of Common Knowledge." *Philosophical Studies* 181:859-877.

Scanlon, Thomas. 2014. *Being Realistic about Reasons*. Oxford: Oxford University Press.

Schelling, Thomas. 1960. *The Strategy of Conflict*. Cambridge: Harvard Business School Press.

Schmid, Hans Bernhard. 2023. *We, Together: The Social Ontology of Us*. Oxford University Press.

Schönherr, Julius. 2021. "Salience reasoning in coordination games." *Synthese* 199(3-4):6601-6620.

Schroeder, Mark. 2007. *Slaves of the Passions*. Oxford: Oxford University Press.

Searle, John. 1990. "Collective Intentions and Actions." In P. Cohen, J. Morgan, and M.E. Pollack, eds., *Intentions in Communication*: 401-415. Cambridge, Mass: Bradford Books, MIT Press.

Setiya, Kieran. 2014. "Love and the Value of Life." *Philosophical Review* 123(3):251-280.

———. 2016. *Practical Knowledge: Selected Essays*. Oxford: Oxford University Press.

Sillari, Giacomo. 2008. "Common Knowledge and Convention." *Topoi* 27:29-39.

Simon, Judith, ed. 2020. *The Routledge Handbook of Trust and Philosophy.* New York: Routledge.

Sinhababu, Neil. 2015. “Advantages of Propositionalism. *Pacific Philosophical Quarterly* 96(1):165-180.

Small, Will. 2012. “Practical Knowledge and the Structure of Action.” In Günter Abel and James Conant, eds., *Rethinking EpistemologyRethinking Epistemology: Volume 2*:133-227. Berlin, Boston: De Gruyter.

Smith, Matthew Noah. 2010. “Reliance.” *Noûs* 44(1):135-157.

———. 2016. “One Dogma of Philosophy of Action.” *Philosophical Studies*, 173:2249-2266.

Stalnaker, Robert. 1978. “Assertion.” In Peter Cole, ed., *Syntax and Semantics 9: Pragmatics*: 315-332. New York: New York Academic Press

———. 2002. “Common Ground.” *Linguistics and Philosophy* 25(5): 701-721.

Strawson, P.F. 1962. “Freedom and Resentment.” In D.F. Pears, ed., *Studies in the Philosophy of Thought*: 71-96. Oxford: Oxford University Press.

Sugden, Robert. 2000. “Team Preferences.” *Economics & Philosophy* 16: 174-204.

———. 2004. “The Logic of Team Reasoning” In Gold, Natalie, ed., *Teamwork*: 181-199. Palgrave-Macmillan.

———. 2015. “Team reasoning and intentional cooperation for mutual benefit.” *Journal of Social Ontology* 1:143-166.

———. 2018. *The community of Advantage*. Oxford: Oxford University Press

Tollefsen, Deborah Perron. 2007. “Group Testimony.” *Social Epistemology* 21(3): 299-311.

Thomas, Kyle A., Descioli, Peter, Haque, Omar Sultan, and Pinker, Stephen. 2014. “The Psychology of Coordination and Common Knowledge.” *Journal of Personality and Social Psychology* 107:657-676.

Thomason, Richard H. 2021. “Common Knowledge, Common Attitudes, and Social Reasoning.” *Bulletin of the Section of Logic* 50: 229-248.

Thompson, Christopher. 2017. “Trust without Reliance.” *Ethical Theory and Moral Practice* 20(3):643-655.

Unnsteinsson, Elmar. 2019. “Silencing without Convention.” *Pacific Philosophical Quarterly* 100:573-598.

Vanderschraaf, Peter. 1998. “Knowledge, Equilibrium, and Convention.” *Erkenntnis* 49: 337-369.

Velleman, David. 1997. “How to Share An Intention.” *Philosophy and Phenomenological Research* 62: 29-50.

Vermaire, Matthew. 2020. “Retraction and Testimonial Justification: A New Problem for the Assurance View.” *Philosophical Studies* 177(12):3959-3972.

Wanderer, Jeremy. forthcoming. “Clarifying Illocutionary Force” *Inquiry*.

Williams, Bernard. 1979. “Internal and External Reasons.” In Ross Harrison, ed., *Rational Action: Studies in Philosophy and Social Science*: 17-28. Cambridge: Cambridge University Press.

Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.

Zimmerman, Michael J.. 2017. “Prospective Possibilism.” *The Journal of Ethics* 21:117-150.