SECOND-PERSON THOUGHT

by

Bryan Chambliss

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2019

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Bryan Chambliss, titled Second-Person Thought and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Terence Horgan — Date: April 15, 2019

Shaun Nichols — Date: April 15, 2019

Jonathan Weinberg — Date: April 15, 2019

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

We hereby certify that we have read this dissertation prepared under our direction and recommend that it be accepted as fulfilling the dissertation requirement.

Terence Horgan — Date: April 15, 2019
Terence Horgan
Dissertation Committee Co-Chair
Department of Philosophy

Shaun Nichols — Date: April 15, 2019
Shaun Nichols
Dissertation Committee Co-Chair
Department of Philosophy

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of sources is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Bryan Chambliss

ACKNOWLEDGEMENTS

This dissertation began as an idea for a series of papers on the rich content of social perception. As my interest in sociality came to outweigh my focus on perception, the project turned into something else. The result is something that I could never have produced alone, and I am grateful to the many people who have shared the journey with me, and guided me along the way.

First, I'd like to thank the members of my dissertation committee—Terry Horgan, Shaun Nichols and Jonathan Weinberg—for sharing their wisdom and expertise. In particular, I'd like to thank Shaun and Terry for many years of feedback on papers, helpful advice, and for teaching me how to do philosophy.

I'd also like to thank the philosophers who introduced me to the subject, and have done so much to enrich my appreciation of it. Whether or not I realized it at the time, David Buller, Jonathan Dancy, Carl Gillett, Tomis Kapitan, and Michael Tye each left an indelible mark on my philosophical interests and temperament. More recently, Jenann Ismael and Michael McKenna introduced me to a cluster of issues surrounding the self, agency and interaction, which led indirectly to the current project.

For continually reminding me of why philosophy is best done with others, I'd like to thank the faculty and graduate students of the University of Arizona philosophy department. I am especially grateful to the members of my cohort, my housemates, and Terry Horgan's work in progress group for years of interesting conversation and helpful feedback. I thank you for all the beers at Time Market, hours at the gym, backyard conversations, and good times at house parties. I'd also like to thank the audiences, commentators and friends at the conferences where I presented some of this material.

My most important debts have less to do with philosophy. This project couldn't even have been started, much less completed, without the support of my family and friends. Thank you for loving me fiercely. Especially when I was too stubborn to see how much I needed, and too dim to see how much I had. I started graduate school with the vague idea that I was moving to Tucson to be trained up as a philosopher. A few years into the process, I realized that what I'd gotten was something much more important. Without Theresa, I'm not sure why I would have finished this project. You fill me with wonder, inspiration, and love. Without my parents, I'm not sure how I could have done so. For their tireless encouragement and support, this dissertation is dedicated to Britt and Ray Chambliss.

I can't tell one from the other,
Did I find you or you find me?

Talking Heads, "This Must be the Place (Naïve Melody)"

TABLE OF CONTENTS

ABSTRACT

Imagine that as you are sitting down to eat a sandwich, you hear someone issue a command: "Stop that this instant!" Perplexed, you look up to see a stranger glaring at you, and come to believe that they told you to stop eating. But in the midst of your ongoing interaction with this person, your thought about the stranger doesn't seem to represent them as either a mere object or agent. Instead, it seems natural to think of them in a second-person way, as a "you" or an agent *with whom I am interacting*. I argue that during face-to-face interactions with other agents, some of our thoughts represent these agents in an irreducibly second-person way. My dissertation defends an account of these irreducibly second-person thoughts.

Chapter 2 introduces second-person thought. I argue that irreducibly second-person thought employs an irreducibly second-person mode of presentation, and that this mode of presentation should itself be understood in terms of its function: recognizing its object as an agent with whom I am engaged in interaction. This recognition-based account is superior to competing accounts that model the second-person mode of presentation in terms of paradigmatic interactions, like communicative interactions (e.g., linguistic exchanges which employ the second-person pronoun 'you'), or cooperative interactions.

Chapters 3 and 4 defend a basic argument for the irreducibility of second-person thought. Among others, John Perry and David Lewis argue that irreducibly first-person thought—variously called essentially indexical thought or *de se* thought—plays an essential role in the explanation of action and yields a distinctive form of self-knowledge. While prominent defenses of the irreducibility of such thought extend only to self-directed thought, chapter 3 develops an analogous argument for thought about other agents, contending that irreducibly *second-person* thought plays an analogous role in the explanation of interaction and yields a distinctive form of knowledge of others. Thus, the same kinds of reasons that have driven many philosophers to accept irreducibly first-person thought can be expanded to give structurally identical arguments for irreducibly second-person thought. Chapter 4 then defends a basic argument for irreducibly second-person thought. Perry and Lewis have taught us that irreducibly first-person thought exists. But the arguments for irreducibly first and second-person thought stand and fall together, so if irreducibly first-person though exists, then irreducibly second-person thought exists too. Thus, irreducibly second-person thought exists.

Having established that irreducibly second-person thought exists, Chapter 5 develops an account of it. Extending the account of Francois Recanati, my account of second-person thought captures the cognitive significance of the second-person mode of presentation as a "mental file" containing distinctively second-person information. The mental files linked to second-person thoughts are populated with information by what Recanati calls "epistemically-rewarding relations" that hold between the thought's thinker and the thought's object. Perceptual acquaintance is the paradigmatic epistemically-rewarding relation, but, in irreducibly second-person thought, the distinctively second-person information results from the thinker's being *directly engaged* with the agent with whom they are interacting. Finally, the referent of a second-person thought is determined not by satisfying the information in the mental file (which could be inaccurate), but by bearing the epistemically rewarding relation to the thought's thinker.

This dissertation argues that while philosophers have rightly recognized irreducibly first-person thought as a distinctive form of thought, they have incorrectly restricted this irreducibility to the first-person perspective. Instead, analogous considerations show that both first- and second-person thought are irreducible. The result is that thought about other agents is different when interacting with them, as opposed to merely observing them. This capacity to recognize an agent as a partner in interaction is characteristic of a distinctively social form of thought, which lies at the root of interpersonal morality.

Chapter 1: Putting Perspective into Social Cognition

I always know where I am. Even when I can't pick out my location on a map, or when I am unsure of where Bryan Chambliss is, I nonetheless know where *I* am. The same goes for who and when. *I* am *here* and *now*. This seems like an unremarkable ability, but we can imagine beings who lack it. Much like creatures whose memories are so vivid as to be indistinguishable from current experiences might lose track of the time of their thoughts, creatures whose experience is too objective, representing only a list of facts, might be unable to place themselves at a specific location. Our ability to *self-locate* is shared by any agent capable of self-conscious thought. Self-conscious thoughts are about the thinker herself, but not by happenstance. In self-conscious thought, one represents oneself in a distinctive first-person manner: representing oneself as oneself. Such thoughts are paradigmatically expressed in English by using first-person pronouns, like 'I', 'me' and 'my'.

My thoughts about everything else seem quite different. For any object that is not me myself, it seems that I might lack knowledge of who, where, or when it is. You are no exception. Though it poses no obstacle to my thinking about you, or to my using the second-person pronoun 'you' to successfully refer to you, I might not know who, where, or even when you are. In this way, my thought about you seems less like self-conscious thought, and more like my thought about other objects in the external world. Whether we think about external world objects in a *de re* manner (*She* is wearing a hat), or by employing something akin to a description (e.g., *The tallest woman in Australia* is wearing a hat), we seem to think of these other objects in an objective, or third-personal, manner.

Combined, these differences suggests a dichotomy between two ways that we might think about something. There is a distinctively first-person form of self-conscious thought, through which we might think about ourselves (and only ourselves!). And there is a third-person form of thought, through which we might think about other objects, or ourselves (when we don't think of ourselves as ourselves). But the dichotomy appears to be exhaustive, as "second-person thought" collapses into a form of third-person thought, albeit third-person thought that we might aptly express using a second-person pronoun. Because this dichotomy seems to wall off the self (me myself!) from the rest of the world, both epistemically and metaphysically, I'll call it the *Subject-Object Wedge* (or S-O Wedge).

The S-O Wedge plays an important role in orienting philosophical work. Is offers philosophers a basic picture of how things are: selves intuitively seem to be metaphysically different from the rest of the world, just as self-conscious thought seems to have distinctive epistemic features. Some of the most perplexing philosophical issues in metaphysics, epistemology and the philosophy of mind and language, hinge upon developing an adequate understanding of the "first-person perspective" or the subjectivity of self-conscious thought, and its relation to the objective or third-person perspective. Thomas Nagel even suggests that most central philosophical issues are aspects of this single problem: the problem of "how to combine the perspective of a particular person inside the world with an objective view of that same world, the person and his viewpoint included" (Nagel 1986, p. 3). In the face of this "world-knot," philosophers have proposed a host of unsatisfactory reductions of one side to the other, or counseled equally unhappy despair, in the form of eliminativism or irreconcilable pluralism.

The underlying ambition of this dissertation is to begin dissolving the hold of the S-O Wedge upon the philosophical imagination. But the first step I propose won't come as unqualified good news. I won't be arguing that the problems of reconciling a first-person perspective with an objective world are illusory or readily resolved. Instead, my aim is to demonstrate that the problem of perspective is more complicated than the S-O Wedge portrays. In short, the problem of perspectivality outstrips the problem of subjectivity, and the latter is best understood as one instance, but not the only instance, of the former. Thus, while I grant that thought takes distinctly first- and third-person forms, and that distinctive philosophical problems are posed by relating these two kinds of thought, I deny that the dichotomy between the first- and third-person is exhaustive. In addition to distinctively first- and third-person thought, one might have distinctively second-person thoughts agents with whom one is engaged in interaction. In the context of an ongoing reciprocal interaction, e.g., one in which you and I seek to coordinate our actions in light of what the other is doing, I seem to know who, where and when you are in a way that mirrors my self-locating knowledge of myself. Thus, my *other-locating* thoughts are about an agent with whom I am interacting, but not by happenstance. Second-person thoughts represent their object via a distinctively second-person mode of presentation—viz. *an agent with whom I am interacting*.

The main contention of this dissertation is that second-person thought is *irreducible*; e.g., that the second-person mode of presentation constitutive of second-person thought cannot be replaced by some combination of first- or third-person modes of presentation. To argue for this claim, I articulate a particular understanding of the second-person mode of presentation, and offer an account of irreducibly second-person thought that assimilates it into the mental files framework that has been defended and used by other philosophers and cognitive scientists. The result is a clarified vision of the role of perspective in thought, and in particular, the role of the second-person perspective in social cognition. The remainder of this introduction frames the project, and draws out some connections between the second-person perspective and existing philosophical work on thought and language.

1.1 Thinking about Other Agents.

Accounts of the second-person perspective are ultimately rooted in the philosophical problems posed by other minds. This might seem like poor philosophical soil to work, as "the problem of other minds" is decidedly unfashionable these days. Despite the interesting work on mindreading and social cognition done in the last 50 or so years, by both philosophers and cognitive scientists, most philosophers simply don't take other minds to pose any distinctive philosophical problem. Jerry Fodor put the sentiment nicely when he noted that the problem of other minds was once a core topic in the philosophy of mind, but that lately, "It's gotten harder to believe that there is a special problem about the knowledge of other minds (as opposed to other anything elses)…" (Fodor 1994, p. 292). The phenomenon that Fodor identifies stems from the fact that there are multiple problems posed by other minds, and that the two most familiar really aren't distinctive and interesting. Accordingly, before discussing second-person thought directly, I'll begin by differentiating some problems of other minds, and noting two further problems that seem to be both interesting and distinctive.[1]

The *Traditional Problem* of other minds is an epistemic problem, which takes the form of a skeptical worry about our knowledge of other minds. Much as it seems possible that one is radically in error about the existence of the external world, it seems possible that one might be radically wrong about the existence of other minds, e.g. I can conceive of a world, phenomenally indiscriminable from the actual one, in which mindless automatons look, sound, smell, etc. exactly like the minded human

[1] I'll largely put aside the question of whether the ensuing problems are interesting, distinctive of other minds, and *genuinely philosophical*. While some clearly are, I confess to finding myself unable to get very interested in the matter of which problems of other minds count as philosophy, as opposed to philosophy-adjacent.

beings of the actual world. If knowledge of other minds involves ruling out such possibilities, then we lack knowledge of other minds. So understood, nothing seems to hinge upon details specific to other minds: this "problem of other minds" is a specific instance of a general skeptical argument. There is a genuine philosophical problem here, but of the sort that routinely exercises epistemologists, not just philosophers of mind. Insofar as the traditional problem is a general skeptical problem, Fodor was right, it isn't a problem distinctive to other minds.

The traditional problem, however, isn't the only one. The next most visible problem of other minds is the *Conceptual Problem* of other minds. This problem seeks an account of how it is possible for us to even think about other minds, much less have knowledge of them, and is thus prior to the traditional epistemic problem. The conceptual problem hinges upon the nature of our mental state concepts, like BELIEF, DESIRE, PAIN, etc., which (one assumes) ought to be generally applicable to both ourselves and to others. Assuming that the meaning of mental state concepts must be tied to their conditions of verification, a problem ensues, as our means of verifying claims about types of our own mental states are quite different than our means of verifying claims about the very same types of mental states of others.[2] The problem, then, invites us to give up either the generality of mental state concepts, the connection between meaning and verifiability, or admit that our mental state thought and talk is not in good philosophical standing. While this problem received extensive discussion in the mid-20th century, most philosophers now reject the postulated connection between meaning and verification. Thus, insofar as the "problem of other minds" hinges upon an outdated account of meaning, it isn't an interesting problem.

Were these the only two philosophical problem posed by other minds, Fodorian pessimism would seem warranted. But they aren't. The epistemic malaise generated by other minds is equally well formulated as a problem concerning the source(s) of our knowledge of other minds, which I'll call the *Source Problem* of other minds. Gomes (2013) formulates this problem as an antinomy:

> **(S1)** If you know that p, then there is a means by which you know that p.
> **(S2)** There is no means by which one can know truths about another's mind.
> **(S3)** We know truths about another's mind.

[2] Wittgenstein and P.F. Strawson crafted influential statements of the conceptual problem of other minds. For more recent discussions of the conceptual problem, see Avramides (2001) and Gomes (2011).

The skeptic about other minds invites us to accept (S1) and (S2), and in so-doing, reject (S3). Bracketing issues related specifically to self-knowledge,[3] (S1) seems fairly unobjectionable. After all, even if knowledge is basic, it isn't mysterious, and if one knows that p, there must be some means by which one knows that p.

Most non-skeptical responses to the problem seek to deny (S2). Traditionally acknowledged sources of knowledge include perception, inference (or reason), testimony, memory, and introspection (or consciousness). Why think that none of these source will afford us knowledge of other minds? Introspection and memory won't work. While introspection is (allegedly) the primary manner in might come to know our own minds, it is by definition not a means by which we can come to know about other minds. Memory appears to be similarly limited to our own mental lives. Testimony won't work either. While testimony is surely one means by which we come to have knowledge of other minds, it isn't the only one.[4] It looks like the root source of our knowledge of other minds must lie in either our perceptual knowledge of other minds, or knowledge by inference.

On its face, perception of other minds isn't an appealing option. We might perceive the bodies, faces and movements that are causally linked to minds, but we don't perceive the minds themselves.[5] Thus, many solutions to the epistemic problem of other minds typically seek to characterize the inferences that generate our beliefs about other minds.[6] But it isn't obvious that we fare better here. Deductive inferences don't seem promising, as it isn't clear from where we get the first principles. Neither do inductive inferences (which seem either to fail to confirm knowledge, or to be based upon a single-case induction). Abductive inference might fare better, but complications abound. While each of these positions has been endorsed historically, and while most philosophers agree that one of them works,

[3] One might worry that there isn't a single way in which we come to know our own minds. For example, Carruthers 2011 argues that our means of access to sensory states is quite different than our access to non-sensory states (especially our own propositional attitudes).

[4] Further, some philosophers, e.g., Stroud 1989 and Cassam 2007, allege that a problem with circularity looms. Surely we must have knowledge of other minds in order to accept testimony about them. If correct, this buttresses the (intuitively true) claim that testimony cannot be the root source of our knowledge of other minds.

[5] This claim has been repeatedly challenged, and various philosophers have argued that our knowledge of other minds is perceptual in nature. For earlier examples, see Dretske 1973, Wittgenstein 1980, and McDowell 1983. The claim has received extensive attention in the last 10 years.

[6] To clarify: Reason is typically taken to be a basic source of knowledge. Reason encompasses both rational intuitions (which have a priori justification) and inferences (which infer a conclusion on the basis of some inference base). Assuming that we don't have a priori knowledge of other minds, I've jumped right into a discussion of inference.

there isn't anything like a current consensus about which of them is the best way to defend inference as a source of knowledge about other minds.

By this point in the dialectic, it should be obvious that one cannot address the epistemic issues raised by other minds without endorsing both some epistemic claims (e.g. an account of what justification consists in, a taxonomy of epistemic sources, etc.) and some basic aspects of a descriptive account of how it is that we form believes about other minds. Only then might one see how our (putative) knowledge of other minds measures up to the epistemic desiderata. It should therefore be unsurprising that philosophical and empirical work on other minds have substantively merged in the second half of the 20th century, and that a large body of work has been produced by both philosophers and cognitive psychologists. In lieu of a literature review,[7] I'll specify some standard terminology that I'll be using to characterize the psychological capacities which underpin our knowledge of other minds.

Knowledge of Other Minds: Some Terminology.

Social Cognition: The cognitive processes that drive our interactions with other agents. This is a heterogenous mix of cognitive processes, including mental states attribution, as well as social attention and memory, social categorization, and affect.

Folk Psychology: The cognitive capacities used in making sense of an agent's behavior. May or may not involve attributing mental states to the other agent for the purposes of predicting and explaining their behavior.

Mindreading (or **Theory of Mind**): The cognitive capacities used in attributing mental states in order to make sense of an agent's behavior. This involves attributing beliefs, desires, intentions, emotions, etc. to other agents. The function of these attributions is typically to predict and explain behavior.

Standard Accounts of Social Cognition: Understand social cognition largely in terms of mindreading. Our interactions with other agents are driven by our abilities to make sense of their behavior, and respond accordingly. Our ability to make sense of other's behavior stems from our mindreading abilities, and paradigmatically, our ability to attribute beliefs to other agents.

There is substantive disagreement about the mechanisms and capacities by which we achieve our understanding of other minds. The *Descriptive Problem* of other minds seeks an account of how it is that we come to understand the minds of other agents, such that we might come to have knowledge of them. This kind of understanding is standardly taken to consist (mostly) in attributing mental states to

[7] Even restricting oneself to mindreading, there is a lot of literature to review, and several good reviews are already available. For examples, see Goldman (2006), Maraffa (2011), and Nichols (unpub).

other agents in order to explain and predict their actions. But the means by which we come to understand other minds is unlikely to be exhausted by mentalizing. Other processes that recognize agents as such, attribute character traits, etc. are likely to be involved. Thus, the Descriptive Problem seeks an account of how it is that we form beliefs about other minds, but also how mindreading is related to the other processes involved in social cognition. If this seems like a purely psychological problem, as opposed to a philosophical one, it isn't. There are at least two kinds of questions that arise when assessing solutions to the Descriptive Problem:

(*How Possibly*) How is it possible to understand other minds?
(*How Actually*) How do we actually go about understanding other minds?

Both of these are live questions. To many philosophers, it isn't obvious how mindreading is possible. The Conceptual Problem called into question the suitability of our mental states concepts for the purposes of attributing mental states to others (or ourselves). More recently, philosophers have argued that understanding others by attributing mental states to them results in a kind of computational intractability.[8] Additionally, there are radical critiques, which contend that mindreading cannot consist in attributing beliefs to others. These once came primarily from radical simulationists (e.g. Robert Gordon, Jane Heal), but more recently have come from those who reject the representational theory of mind in favor of some other 4E (embodied, embedded, extended, enactive, etc.) account of the mind. These challenges contend that mindreading, as standardly construed, isn't an adequate source of knowledge about other minds.

Of course, not all of these challenges contend that mindreading couldn't be attribution-based, or realized by computations over representations, etc. Some merely contend that, for one reason or another, the best explanation lies elsewhere. But on either the how-possibly and the how-actually versions of the problem, it is clear that philosophers themselves haven't been content to view the descriptive problem of other minds as unphilosophical. Even above and beyond its bearing on the epistemic issues raised by other minds (like the source problem), the descriptive problem poses an interesting and distinctive philosophical problem about other minds.

On balance, then, the problem of other minds is hardly dead. The Source Problem and the Descriptive Problem show that other minds pose interesting and distinctive philosophical problems. These

[8] See Hutto (2004) and Zawidzki (2013) for examples of this argument.

problems raise widely recognized issues for both philosophers (e.g., epistemologists, philosophers of cognitive science) and for cognitive scientists (e.g., developmental psychologists, social neuroscientists). And while the primary interest of each group is often on different aspects of the problem, philosophers and cognitive scientists are jointly engaged in addressing some aspect of the problem of other minds. But existing research programs investigate and explain other minds in quite different ways, and disagreements between rival approaches often implicate both philosophical and empirical issues. It is in such disputes that discussions of "the second-person perspective" have taken root. To illustrate the potential importance of the second-person perspective, consider what is sometimes called the "Interactive Turn" in social cognition research.

1.2 Turning to Interaction? The Interactionist Critique of Standard Accounts of Mindreading.

The "Interactive Turn" consists of the renewed focus upon social interaction that has been found in some research on social cognition. Although positive accounts have been offered, this "turn" is predominantly critical, arguing that standard accounts of social cognition lack the resources to explain even ordinary interactions. In particular, it is alleged that standard accounts of social cognition cannot explain the kinds of meaningful, face-to-face interactions that occur in people's daily lives. Instead, mindreading must be dethroned as the central element of social cognition, and interaction itself must occupy a more central role in accounts of social cognition. To better appreciate the critique, what it does and doesn't accomplish, we must further unpack the idea of a standard account of social cognition.

We naturally understand other agents as having mental states, and we use this understanding to guide our interactions with them. When someone's behavior perplexes us, we do this consciously—what could he be thinking?—as we puzzle out people's motivations and actions. But belief-desire psychology needn't be deployed consciously, nor merely as a stance that agents take up to interpret others. In standard accounts, belief-desire psychology operates akin to a high-level description of social cognition itself. Social cognition consists centrally of mindreading, e.g., capacities for attributing mental states (beliefs, desires, intentions, etc.) to others in order to predict and explain their actions. Thus, standard accounts of social cognition privilege both a set of capacities (mindreading) and a function (predicting and explaining action) in the psychological explanation of interaction. Among proponents of such accounts, the nature of these capacities remains debated (is mindreading more like deploying a theory, running a simulation, or both?), but however it is done, it is agreed that

mindreading occurs pervasively in ordinary social interactions. Such accounts endorse what Shannon Spaulding calls the *broad scope of mindreading claim*, "that mindreading is the primary, pervasive way we understand others, that mindreading is essential for navigating the social world, and that we could not understand and interact in our social environment as successfully and easily as we do if we were not avid mindreaders" (Spaulding 2018, pp. 8-9).

Standard accounts are also *observationally inclined*. Interacting with someone and merely observing them seem quite different. This is born out not only as a fact about our experiences of each, but by experimental evidence concerning the nature of the processes involved and their development.[9] But mindreading itself doesn't seem to mark a difference between interaction and mere observation: one can attribute mental states to others in order to predict and explain their actions equally well in either condition. Further, putting aside the case of first-person mindreading (mindreading oneself), mindreading is canonically taken to be third-personal.[10] Thus, standard accounts are often faulted for being too "intellectualist," e.g. for collapsing social cognition into behavioral observation and inference. And while it isn't at all obvious that standard accounts must be so observationally inclined, it is clear enough that they have been. This occurs both as a presupposition about the nature of mindreading, and a methodological stricture about how it should be studied. For example, the neural mechanisms of mindreading have typically been studied by having subjects read vignettes that describe social interactions, or by watching videos of social interactions. While it is perfectly reasonable to study mental state attribution with such stimuli, the extent to which real-life interaction recruits processes of attribution so understood remains an open question. Standard accounts have been quick to assume that they do, and critics question this assumption. It is at this juncture that discussions of the second-person perspective— "second person interaction" (Gallagher 2001), "second person context" (Hutto 2004), "second person stance" (Ratcliffe 2006), "second-person relationship", "second-person access" (Zahavi 2007), and "second person perspective" (Gallagher 2009, Schilbach 2010)—have worked their way into the philosophical and scientific literature on other minds.

A renewed focus upon interaction thus furnishes an *Interactionist Critique* of standard accounts of social cognition. This critique is typically portrayed as radical in nature, claiming that standard accounts offer

[9] For reviews, see Schilbach et. al. 2013 and Reddy 2008.

[10] Perusing literature on the nature of mindreading, one can really see the S-O wedge at work. In the context of this dissertation, this serves as a reminder that not only are descriptive problems of other minds genuinely philosophical problems, but that blatantly philosophical presuppositions inform solutions to the descriptive problem.

a fundamentally mistaken account of how individuals understand each other. The alleged mistake is that standard accounts, by endorsing the broad scope of mindreading alongside its observational inclinations, have failed to take interaction seriously. This failure manifests in a variety of ways, many of which center not merely on issues specific to social cognition, but on the nature of cognition more generally. Thus, articulations of the interactionist critique are often given not merely on the way to a rejection of the standard account of social cognition, but in the service of embedded, embodied, extended of enactive accounts of cognition more generally. For example, De Jaegher, Di Paolo and Gallagher 2010 argue that we can't make sense of some interactions in terms of two cognizers who individually represent facets of the world, and attempt to predict and react to each other's behavior. Instead, in some cases, interaction literally constitutes social cognition. They argue that "interactive processes are more than a context for social cognition: they can complement and even replace individual mechanisms" (De Jaegher, Di Paolo and Gallagher 2010, p. 441). In their hands, the relationship between social interaction and social cognition is used to motivate their general skepticism about the representational theory of mind, and social cognition is used to illustrate the more general need for an enactivist account of cognition.

Nor are De Jaeger, Di Paolo and Gallagher unique in hanging general issues about cognition upon discussions of social cognition. The result of this tendency to hang is that statements of the interactionist critique often hinge upon rejecting some very general claim about the foundations of the cognitive sciences—representationalism, the computational theory of mind, the relationship between cognition and experience, etc. Thus the "interactionist challenge" is often presented as a skeptical challenge to standard representational/computational accounts of cognition. As such, it is unsurprising that the challenge itself has been subject to trenchant criticism,[11] and that the critique itself is often set aside. Many philosophers of cognitive science are hesitant to take evidence concerning the details of some specific cognitive capacity to support such sweeping theoretical reform. It's like hitting a thumbtack with a mallet; the tool seems incommensurate to the job.

I too reject the radicalism built in to the "interactionist critique," but I'm not convinced that we can simply set the worries about interaction aside. Many formulations of the interactionist challenge run together two separate issues: concerns about the relation between interaction and social cognition,

[11] For examples, see Overgaard and Michael 2015, Spaulding 2015, Schonherr and Westra 2019.

and skepticism about the representational or computational theory of mind. Of course, if there really is a compelling case to be made for skepticism, then this challenge must be answered. But formulating the interactionist challenge in this manner runs the risk of throwing out the baby (the genuinely interesting issue about interaction) with the bathwater (the more general issues about representation and computation). Instead, the key to finding a better, more interesting, version of the interactionist challenge is to decouple these concerns. By putting aside the skepticism about the role of representation and computation in social cognition, we can see the interactionist challenge as issuing a call for a kind of pluralistic account of social cognition.[12]

This second, non-skeptical, reading of the interactionist challenge merits a response from accounts of social cognition. While the evidence that interactionists appeal to might not support a revolutionary change to the foundations of the cognitive sciences, it does pose problems for standard accounts of social cognition. In short, it suggests that social cognition really is different in the context of meaningful, face to face interaction than in a merely observational context. This point is typically granted, shortly before pointing out that the difference amounts only to a difference in affective processing.[13] But I don't find this approach attractive—it puts heavy weight on explaining aspects of interactions in terms of affective processing. While I don't deny that there are affective differences, I see no reason to think that these exhaust the differences. Whether or not our affective processing is different in the context of interaction, one wants to claim that something about the thoughts themselves are different too. Specifically, there seems to be something different about how we understand others in merely observational contexts and in the context of an interaction.

It is at this juncture that an account of second-person thought can do real work. When we are merely observing another agent, we have third-person thoughts about them. Social cognition is genuinely different in the context of interaction, because in this context, we might have second-person thoughts about other agents. Thus, second-person thought, if there is any such thing, marks a genuine difference between interaction and mere observation. But it need not carry radical implications. So long as one

[12] The interesting challenge here comes from *social interaction* and not merely interaction. The "interactionist turn" ought to be just as keen to place sociality at the center of an account of social cognition, as it is to place interaction at the center. Interaction seemed key to the 4E crowd because their beef was really with traditional (representational, computational) accounts of cognition. But, as Reddy (2018) points out, it is actually sociality that seems to matter. If social interaction can be understood in more traditional terms, then so much the better. Reddy herself doesn't think it can, but I do.

[13] Separated from their more radical trappings, this is basically the claim of Reddy 2008, 2018 and the suggestion of Schilbach et. al. 2013.

can develop a suitable account of second-person thought, nothing about this approach requires a retooling of the foundations of the cognitive sciences.

1.3 A Response: Towards the Second-Person Perspective.

It is one thing to claim that an account of the second-person perspective would be philosophically useful. It is quite another to develop such an account. This is particularly pressing, as many of the few philosophers who have devoted much attention to second-person phenomena have rejected it as incoherent. In a discussion of the senses attached to demonstratives, Richard Heck denies that:

> "…there is any such thing as an essentially indexical second-person belief. The phenomenon of the second-person is a linguistic one, bound up with the fact that utterances, as we make them, are typically directed to people, not just to the cosmos…The word 'you' has no correlate at the level of thought." (Heck 2002, p. 12)

Kent Bach concurs, adding that for the second-person pronoun 'you', "although there is a linguistic rule that determines its reference, relative to a context, as the addressee, there is no corresponding second-person way of thinking of someone" (Bach 2007, p. 416). In each case, the second-person mode of presentation might correspond to a stance that one takes up, but while there might be distinctively first-person modes of presentation, the second-person perspective has no place at the level of thought.

What philosophical resources might one turn to when developing an account of the second-person? Amongst contemporary analytic philosophers, the most visible discussion of the second-person perspective concerns the *second-person stance* that agents take up when they address one another, or take themselves to be apt targets for such address. For example, while discussing moral address, Stephen Darwall defines the second-person standpoint as "the perspective you and I take up when we make and acknowledge claims on one another's conduct and will" (Darwall 2006, p. 3).[14] When we address moral claims to each other, e.g. as demands, requests, reproaches, etc., we take up the second-person stance. But address performed from the second-person stance is a norm-governed activity, and aptly

[14] A precautionary note on terminology: A glance at the literature reveals a number of things that follow the phrase "second-person": perspectives, viewpoints, point of views, stances, standpoints, modes of presentation, thoughts, etc. Sometimes nothing follows the phrase, suggesting that although the modifier 'second-person' is important, what is being modified isn't. To avoid the ensuing confusion, I stipulate jargon in Chapter 2. On my usage, the second-person *perspective* (or *viewpoint*) is a generic term which applies to both stances that agents might take up, or to ways that a thought might represent its object. The second-person *stance* (or *standpoint*) is a stance that might be taken up by an agent, and the second-person *mode of presentation* (or the second-person perspective in thought) is a way that a thought represents its object.

taking up the second-person stance presupposes that the "addresser" and the "addressee" meet certain conditions. First, it presupposes that the addresser has a kind of standing or authority to address the addressee in question. Second, it presupposes that the address in question furnishes the addressee with distinctive kind reason to comply (a second-person reason). Finally, it presupposes that the addressee is answerable or accountable (in some way) to the addresser. Darwall contends that many central moral notions (like obligation, demand, responsibility, etc.) can only be understood in the framework provided by the second-person stance.

Other highly visible discussions of the second-person perspective are found in the phenomenological tradition, and amongst certain developmental psychologists and social neuroscientists. I've canvassed some of these sources in the previous section. A wide range of phenomenologists have given accounts of *intersubjectivity*—which means something like the "engagement between subjectivities" that takes place in interactions between agents, and which gives rise to interpersonal understanding (when things go well). I won't hazard a more rigorous definition here, or even attempt to sort out the various ways in which 'intersubjectivity' gets used.[15] Instead, I'll note that accounts of the phenomenology of interaction often appeal to the second-person perspective. This typically takes the form a linguistic metaphor, in the form of the claim that we experience other people as a you (or a thou), or the claim that interaction has a dialogical structure. While the role of the second-person perspective is explicitly acknowledged in such accounts, this feature of experience is alleged to be something that standard accounts of social cognition miss. Thus, as we saw in the previous section, radical claims naturally take root in accounts of the phenomenology of interaction: mindreading is unlike our fundamental means of understanding other agents, which are perception-like, fundamentally enactive, etc.

Similarly, developmental psychologists and social neuroscientists have found distinctive roles for the second-person perspective. This might take the form of claims about infant perceptual capacities, as when Gopnik, Metlzoff and Kuhl tell us that "quite literally from the moment we first see other people, we see them as people…To see someone as a person is to see a face, not a mask; a "thou" not an "it"" (Gopnik, Meltzoff, and Kuhl 2001 p. 24). Alternatively, it might take the forms of a claim about the mature capacity, as when Schilbach et. al. tell us that "social cognition is fundamentally different when we are in interaction with other rather than merely observing them" (Schilbach et. al.

[15] See Crossley 1996, Gillespie and Cornish 2009 and Zahavi and Overgaard 2013 for useful discussion.

2013, p. 393). Both of these are instances of what Vasudevi Reddy calls a "second-person approach to knowing other minds," which highlight the role played by the interaction between the agents. For example, Reddy argues that "in order to understand how infants develop an awareness of other minds, psychology needs to pay attention to what they do in engagement with other people" and that "both developmentally and experientially, recognizing and being recognized by a *You* is primary for understanding other people" (Reddy 2008, pp. 232-3).

These is not an exhaustive list of philosophers and psychologists utilizing concepts related to the second-person perspective. Rather, I've gestured towards three of the most visible instances of philosophers and cognitive scientists arguing that, in some sense, the second-person perspective is indispensable. The differences between these discussions are stark. When each makes a claim about "the second-person perspective," they make claims about quite different things: stances are things that agents take up, experiences are things that they have, and cognitive and neural processes are subsystems of agents. Further, although each claims that the second-person perspective is indispensable, its indispensability lies in accomplishing quite different tasks. Nonetheless, there is enough commonality between the three discussions to furnish an intuitive starting point: the second-person perspective is that of a participant in an interaction, as opposed to a mere observer.

To get traction upon this intuitive idea, I suggest that we first turn to something more familiar: the nature of self-conscious thought. Self-conscious thought not only has a distinctive referent (the thinker of the thought), but this referent is thought of in a distinctive manner. In self-conscious thought, one thinks about oneself as oneself. Such thoughts are typically expressed by using a distinctive set of first-person pronouns, like 'I', or 'me myself', and are widely thought to have distinctive epistemic and explanatory features.

Further, philosophers have argued for the irreducibility of these first-person thoughts, and these arguments have been scrutinized,[16] and widely endorsed. It is by attending to the case for irreducibly first-person thought, and then considering whether or not an analogous case can be made for irreducibly second-person thought, that I intend to develop a case for the latter. I'll begin with a set of arguments for irreducibly first-person thought drawn from what I'll call the *Castaneda-Perry-Lewis*

[16] Though maybe not as thoroughly as they ought to be. See Cappelen and Dever (2013) on the suspiciously wide range of conclusions that are routinely asserted, and then "justified" by citing Lewis (1979) and Perry (1979).

(C-P-L) Tradition. Arguments in this tradition identify the seemingly distinctive explanatory and epistemological features of self-conscious thought, and then use these distinctive features to show that first-person thought cannot be reduced to third-person thought.

To give a well-known example, John Perry has argued that self-locating thought, a kind of first-person thought about who, when or where I am, plays an indispensable role in rationalizing action. Perry's example is familiar:

> ***Sloppy Shopper*:** I once followed a trail of sugar on a supermarket floor, pushing my cart down the aisle on one side of a tall counter and back the aisle on the other, seeking the shopper with the torn sack to tell him he was making a mess. With each trip around the counter, the trail became thicker. But I seemed unable to catch up. Finally it dawned on me. I was the shopper I was trying to catch…I believed at the outset that the shopper with a torn sack was making a mess. And I was right. But I didn't believe that I was making a mess. That seems to be something I came to believe. And when I came to believe that, I stopped following the trail around the count, and rearranged the torn sack in my cart. My change in beliefs seems to explain my change in behavior. (Perry 1979, p. 3).

Perry's argument is straightforward.[17] There is a change in belief that explain Perry's action in cases like Sloppy Shopper. This change in belief only rationalizes Perry's action, but can only be captured in first-person language, as opposed to co-referential third person language. But this is no mere linguistic phenomena, as the change in belief can only be captured in first-person language because the change consists (partly) in distinctively first-personal mental states. While various questions about the conclusion remain open (What about the mental state is first-personal? Which types of mental states are irreducibly first-personal? Etc.), many philosophers have been convinced that irreducibly first-person mental states play an ineliminable role in explaining action.

Alternatively, self-conscious thought also seems to have distinctive epistemic features. When I notice someone else wince in pain, and think that her headache hurts, I have a third-person thought about *her*. But when I notice a twinge of pain and think to myself that my headache hurts, I have a first-person thought about *me myself*. Many philosophers have argued that we each enjoy a kind of privileged access to our own mental life. While we might not have privileged access to our entire mental life, we nonetheless have a manner of accessing our mental life that is not shared by others. Similarly,

[17] I say more about the structure of Perry's argument in chapters 2 and 3. For now, I will simply note that in calling Perry's argument straightforward I am not suggesting that it is always (or even often) construed correctly.

discussions of the epistemology of self-conscious thought often discuss whether or not it has unique epistemic properties. For example, when I think about myself as myself, I might misattribute a property to myself. I might think I am hot, when I am really cold. But there seems to be a sense in which I couldn't be wrong about who my self-conscious thoughts are about. They are about me myself. When a judgment is immune to this kind of error, we say that it has the property IMMUNITY (short for Immunity to Error through Misidentification relative to the first-person pronoun) (Shoemaker 1968, Prosser and Recanati 2012). A judgment of the form "I am F" has IMMUNITY when, given the information on which the judgment is based, it is not susceptible to error about who or what is being judged to be F. Should any judgments have IMMUNITY, it seems that only irreducibly first-person thoughts might have it.

Can an account of irreducibly second-person thought be modeled on existing accounts of self-conscious thought? This dissertation is an extended meditation on how to do so. I also develop analogous arguments for irreducibly second-person thought. Much as the first-person perspective plays a distinctive role in explaining action and in capturing the distinctive epistemic properties of self-conscious thought, the second-person perspective plays a distinctive role in explaining interaction and in capturing the distinctive epistemic properties of our knowledge of other agents in the context of interacting with them.[18] Such claims about explaining interaction extend an existing philosophical literature on self-conscious thought, but also arise naturally amidst work in the cognitive sciences on cognitive development and social cognition. If correct, this shows that the S-O wedge is, in part, self-defeating. The very argumentative resources that have convinced many philosophers of irreducibly first-person thought, can be mobilized to argue for irreducibly second-person thought as well. In this sense, there is nothing special about the irreducibility of the first-person perspective, as it is an instance of the irreducibility of perspective more generally. Moreover, it puts the first- and second-person perspective on equal footing—while they certainly don't have identical features, the case for their irreducibility seems to stand and fall together.

[18] Parallel claims about the epistemic properties of first- and second-person thought are tricky to motivate intuitively, as many philosophers have intuitions to the effect that the only epistemically salient distinction in the neighborhood is between knowledge of self and knowledge of others. Consequently, I won't base my case for irreducibly second-person thought upon such considerations here. Needless to say, this isn't because I grant that there is some epistemic asymmetry between the first- and second-person perspectives. I'm just picking the easier fight.

1.4 Outline of the Dissertation.

I've laid out the what and the why of this project: I'm going to argue for an account of irreducibly second-person thought, in part to respond to issues in the philosophy of cognitive science that stem from the problem of other minds (broadly construed). I've also gestured at the resources that I intend to deploy when developing this account: I aim to draw upon accounts of singular thought and the irreducibly first-person perspective. What remains is to sketch the path that I will follow in the following chapters. The case for irreducibly second-person thought is developed in the following fashion.

Chapter 2 identifies the core topic of the dissertation: second-person thought. Second-person thought represents its object via a second-person mode of presentation. More generally, the second-person perspective is intimately related to interaction, and existing accounts understand the second-person mode of presentation in terms of linguistic integrations or cooperative interactions. I argue that both kinds of accounts are too restrictive—by focusing on one particular type of interaction (linguistic, cooperative) they over-constrain second-person thought. Instead, I develop an account of the second-person mode of presentation by focusing on the function that it plays in interaction. The second-person mode of presentation should itself be understood in terms of its function: recognizing its object as an agent with whom I am engaged in interaction. This recognition-based account is superior to competing accounts that model the second-person mode of presentation in terms of other functions like mutual awareness. One important upshot of this chapter is that its conclusion allows one to distinguish the feature of second-person thought from the features of the second-person language that one might (or might not!) use to express such thoughts.

Chapters 3 and 4 defend a basic argument for irreducibly second-person thought. Among others, Hector-Neri Castaneda, John Perry and David Lewis argue that irreducibly first-person thought—variously called essentially indexical thought or *de se* thought—plays an essential role in the explanation of action and yields a distinctive form of self-knowledge. While prominent defenses of the irreducibility of such thought extend only to self-directed thought, chapter 3 develops an analogous argument for thought about other agents, contending that irreducibly *second-person* thought plays an analogous role in the explanation of interaction and yields a distinctive form of knowledge of others. Thus, the same kinds of reasons that have driven many philosophers to accept irreducibly first-person thought can be expanded to give structurally identical arguments for irreducibly second-person

thought. I take this to support the following conditional claim: if the irreducibly of first-person thought is shown by the kinds of arguments offered by the CPL tradition, then the irreducibly of second-person thought is shown by analogous arguments.

Chapter 4 defends a basic argument for irreducibly second-person thought. Castaneda, Perry and Lewis have taught us that irreducibly first-person thought exists. But the arguments for irreducibly first and second-person thought stand and fall together, so if irreducibly first-person though exists, then irreducibly second-person thought exists too. Thus, irreducibly second-person thought exists. Chapter 4 examines some skeptical worries that have been voiced concerning irreducibly first-personal thought. For example, Herman Cappelen and Joshua Dever (2013) have argued that the classic arguments for irreducibly first-person thought fail badly. They contend that many of the famous examples of self-locating thought (like Perry's example of the Sloppy Shopper) are nothing more than Frege cases which establish a general type of opacity involving action explanations. My response to this argument seeks not only to show why Perry's argument works, but to clarify its character: Perry is offering a form of high-level empirical argument for irreducibly first-person thought.

Further, it seems fairly clear that we can communicate such thoughts. After all, when John Perry follows the trail of sugar, trying to find the sloppy shopper, he realizes that it is he himself who is making the mess. And he might convey this realization to another shopper—"I was making the mess the whole time!" But on my account of essentially indexical thought, such thoughts represent their objects via a distinctively first-person mode of presentation. Thus a puzzle arises about how such thoughts could be successfully communicated. I deflate this puzzle by sketching an account of how essentially indexical thoughts might be communicated, and in particular, how we express essentially second-person indexical thoughts by using pronouns.

Having established the irreducibly of second-person thought, I turn to a lingering source of skepticism. In an attempt to remain neutral on various issues that surround discussions of irreducibly first- or second-person thought, I articulated a generic account of such thought, which defines such thoughts in terms of modes of presentation. But many philosophers are skeptical about this notion, especially if modes of presentation turn out to be a component of the content of thought.[19] To respond to this

[19] On my account, the irreducibility of second-person thought doesn't entail that irreducibly second-person modes of presentation are thought components. But it isn't inconsistent with that claim about content either.

skepticism, *Chapter 5* develops an account of the second-person mode of presentation that is independently appealing, and (I think) well-positioned to placate such skepticism. Extending the account of Francois Recanati, I offer an account of second-person thought which captures the cognitive significance of the second-person mode of presentation as a "mental file." Mental files are individuated in terms of their function, which itself is articulated by appeal to what Recanati calls the "epistemically-rewarding relations" that hold between the thought's thinker and the thought's object. The epistemically-rewarding relation populates the file with information which is presumed to be about the referent of the file. But, the referent of a mental file is determined not by satisfying the information in the mental file (which could be inaccurate), but by bearing the epistemically rewarding relation to the thought's thinker. Perceptual acquaintance is the paradigmatic epistemically-rewarding relation, but, I articulate a type of mental file (interlocutor file) which accumulates information about an agent with whom one bears the epistemically-rewarding relation of *direct engagement*. I then sketch an account on which the mental files linked to second-person thoughts are interlocutor files. On this account, the "irreducibly second-person mode of presentation" which drives the irreducibility of second-person thought, turns out to be nothing more than a distinctive type of mental file.

The *Conclusion* offers some thoughts about the importance of irreducibly second-person thought. Most basically, the irreducibility of second-person thought shows that we must endorse a kind of pluralism about social cognition. Overall, this shows that while philosophers have rightly recognized irreducibly first-person thought as a distinctive form of thought, they have incorrectly restricted this irreducibility to first-person thought. Instead, analogous considerations show that both first- and second-person thought are irreducible. The result is that thought about other agents is different when interacting with them, as opposed to merely observing them. This capacity to recognize an agent as a partner in interaction is characteristic of a distinctively social form of thought, which lies at the root of interpersonal morality.

Chapter 2: The Second-Person Mode of Presentation

What is the second-person perspective? Most generally, it is the perspective of a participant in an interaction, as opposed to that of a mere observer. The practical stances taken up during moral address and linguistic communication provide two familiar examples. Discussing specifically moral exchanges, Stephen Darwall identifies the second-person standpoint as "the perspective you and I take up when we make and acknowledge claims on one another's conduct and will" (Darwall 2006, p. 3). Thus, when addressing each other morally, or taking one to be an apt target of such address, agents take up the standpoint of a participant in a moral exchange, and not a mere observer to one. Darwall argues that, whether explicitly recognized by the agent or not, the "I-you-me structure of reciprocal address runs throughout thought and speech from the second-person point of view" (*ibid.*, p. 3).

Similarly, certain communicative acts can't be performed by mere observers, and instead require one to take up the standpoint of a participant in a conversation. Rebecca Kukla and Mark Lance discuss vocatives ("Yo, Emma!") and acknowledgements ("I hear you"), speech acts that are "inherently *second-personal*: they are in each case *directed at you*. They not only pick out a concrete target audience, but inherently address themselves to that audience" (Kukla and Lance 2009, p. 135). The commonality between these cases is that in both forms of address, an agent takes up, or adopts, the standpoint of a participant to the interaction. It is from this second-person stance that one both might address another agent, morally or communicatively, and take herself to be an apt target of address.

This paper aims to clarify our understanding of the second-person perspective. But it does so not by further characterizing its normative structure, e.g. by articulating the distinctive set of norms governing the action of agents who occupy the second-person stance, or the distinctive structure of the reasons that agents might acquire upon taking up this stance. Instead, I aim to develop an account of a capacity for thought that is necessary for one to occupy the second-person stance. After all, only certain kinds of things can occupy the second-person standpoint. Cognizers and agents can, rocks and chairs cannot. Further, one cannot occupy the second-person standpoint towards just anything. One can only take up the second-person stance towards other agents, and in particular, other agents with whom takes oneself to be interacting. One might erroneously take up the second-person stance towards non-

agents, or agents that one isn't interacting with. But even in such cases, one can take up the second person stance only towards something that erroneous thinks is an agent, and, specifically, an agent that one is interacting with.[20]

These limitations yield a natural hypothesis about the capacities required to take up the second-person stance: only a thinking agent, capable of thinking of others as being engaged with them in interaction, can take up the second-person stance. What is it to think of another agent as being engaged with me in interaction? Let's call such thoughts second-person thoughts. Second-person thoughts are distinctive in how they represent their objects, which I'll capture by saying that a *second-person thought* is a thought that represents its object via a second-person mode of presentation.

This paper seeks to clarify the second-person mode of presentation while remaining neutral on general issues concerning the metaphysics of modes of presentation.[21] For even putting aside issues concerning what a mode of presentation is, philosophers disagree widely about how to understand the second-person mode of presentation. Some philosophers write about "you-thoughts" which are modeled closely upon the communicative role of the second-person pronoun 'you' (e.g., Heck 2002, Rodl 2007, Longworth 2007). Others understand second-person thought to arise from the common knowledge of joint intentions found in cooperative situations (e.g., Heal 2014). Still others understand second-person thought in terms of mutual awareness, or "interpersonal self-consciousness" (e.g., Peacocke 2014, Salje 2017). I argue that the second-person mode of presentation ought not be articulated by appealing to paradigm types of interaction, such as linguistic interactions or cooperative interactions. Instead, the distinctively second-person mode of presentation should be articulated by appeal to the function that second-person thought performs during interaction. Further, I argue that this function consists in recognizing the object of thought as an agent with whom I am engaged in interaction. Accordingly, I defend a recognition-based account, on which the second-person mode of presentation represents its object as *an agent with whom I am engaged in interaction*.

[20] This way of putting things raises interesting questions about the beginnings and endings of interactions. For example, I might begin a conversation by addressing someone that I'm not yet interacting with (e.g. by yelling "Hi!"). And I might end a conversation with someone who continues to attempt to interact with me (e.g. a salesperson who continues to speak with me after I say "No thank you" and walk away). I won't take up these issues here.

[21] Some philosophers take modes of presentation to be an aspect of the content (e.g. Fregean senses). Others treat them as an aspect of the vehicle of thought (e.g. many proponents of the "mental files" framework). This chapter argues for a gloss on the second-person mode of presentation that might be deployed by proponents of either of these accounts.

My argument for the *Recognition-Based Account* of the second-person mode of presentation draws upon the insights of existing accounts of second-person thought, while clearing away their misleading features. *Section 1* regiments some terminology, and distinguishes question about second-person thought from surrounding issues. Most accounts of second-person thought model the second-person mode of presentation on a paradigmatic kind of interaction. *Section 2* discusses and rejects two such accounts, based on communicative interactions (communication-based accounts) and cooperative interactions (cooperation-based accounts). Because each account focuses on a subset of the interactions that second-person thought occurs in, the resulting accounts are too narrow. This result seems to generalize to any account of the second-person mode of presentation based upon a paradigm type of interaction. Instead, we ought to understand the second-person mode of presentation in terms of its function. *Section 3* examines two such accounts, and develops an account of the second-person mode of presentation based upon the recognition function that second-person thought plays in interaction. I conclude in *Section 4* by clarifying what I have, and have not, offered by way of the recognition-based account.

2.1 The Second-Person Perspective and Interaction.

The second-person perspective is the perspective of a participant in an interaction, as opposed to a mere observer. However, beyond this initial point of agreement, philosophical discussion of "the second-person" is often at cross purposes. To avoid confusion, I'll begin by stipulating some terminology. Let the ***second-person perspective*** be a general term. It applies equally to the manner in which a thought represents its object, the character of one's experience during an interaction, the grammatic features of certain bits of language (e.g. second-person pronouns like 'you'), and the kind of stance that an agent might take up during interaction. It is in this general sense that philosophers agree that the second-person perspective is intimately related to interaction. But this core idea might be pursued and elaborated in different ways.

The ***grammatical second-person*** is the most familiar development of the second-person perspective. Two features of public language that play a role in the apt expression of thoughts are indexicality and the marking of grammatical person. A linguistic expression is an *indexical* when its reference can shift due to changes in the context in which the expression is used. For example, suppose that Kim and Jessie both utter the sentence 'I am tired.' In each case, the referent of 'I' is fixed differently, and Kim and Jessie say two different things—Kim says that she is tired and Jessie says that

he is tired. Paradigmatic examples of indexicals include pronouns like 'I', 'you', 'she', and 'he', demonstratives like 'here', 'now' and 'that', and adverbs like 'today'.

Indexicality is a general property of languages that contain indexicals. The indexicality of a language manifests in various forms of context sensitivity, including different ways of marking the speaker's spatio-temporal frame of reference. For example, in English one linguistically marks spatial and temporal proximity to the speaker. Objects that are close at hand are located "here," while objects that are further removed are over "there." Knowledge about who the speaker is, and their spatial location, is needed to accurately interpret such claims. In this way, the language used by a speaker reflects their "frame of reference."[22]

But the speaker's frame of reference is not exhausted by spatial and temporal proximity. It also includes how the speaker relates to other agents, which is grammatically marked as well. *Grammatical person* marks various relations, typically between a speaker (the first person), those being addressed by the speaker (the second person) and others (the third person). Grammatical person can be marked on nouns, adjectives and verbs, but I will focus on pronouns. In English, the relations to the speaker that are marked by grammatical person can be used to define sets of personal pronouns, such as first-person pronouns (I, we), second-person pronouns (you, thou), and third-person pronouns (he, she, it, they).

One way is to articulate the second-person perspective as a practical stance that agents take up during interaction. Let the ***second-person stance***, or ***second-person standpoint***, be the practical stance taken up by participants in an interaction. Work on the second-person stance might focus on the nature of the acts that participants in an interaction perform, the structure of their reasons for performing them, or on the normative structures that govern such interactions. So understood, the second-person stance is the typical subject of the literatures mentioned earlier on moral address and certain communicative interactions.

[22] Linguists call these forms of indexicality *deixis*. Deictic expressions are those that make use of a deictic center (I, here, now) and linguistically encode the speaker's frame of reference. This includes spatial deixis (contrasting here vs. there), temporal deixis (contrasting now vs. before, or now vs. later) and person deixis (contrasting I vs. you vs. he/she/it).

Although the second-person stance is not the focus of this paper, I've noted that not just anything can take up the second-person stance. I hypothesized that only agents with the right kind of cognitive equipment—namely the capacity for second-person thought—can take up the second-person stance. One might deny this. Maybe the second-person stance merely concerns giving and asking for certain types of reasons, and no explanation in terms of the agent's underlying psychology is required. If so, explanation of the second-person stance simply bottoms out, so to speak, in the normative considerations internal to the stance itself. Alternatively, occupying the second-person stance might be explained in terms of something external to the stance itself, just not in terms of anything that manifests a perspective. For example, if cognition isn't in the business of representing at all,[23] then *a fortiori* the cognition required for occupying the second-person stance doesn't represent its objects via any personal perspective (much less a specifically second-person perspective).

While interesting, these alternatives needn't be ruled out before discussing second-person thought.[24] For claims about the second-person perspective are also naturally enough interpreted as claims about thought. Imagine that a stranger waves in your direction, and taking them to be waving at you, you wave back. As it turns out, that the stranger was waving at someone behind you. Upon realizing this, you realize that you made a mistake. Your thoughts about the stranger seem to be mistaken not in their referent, but in how the thought presented its referent. It seems like participants in an interaction have a distinctive manner of thinking about those with whom they interact, and thoughts cannot manifest the second-person perspective in thought towards things that we don't take to be interacting with us. If so, then thoughts that manifest the second-person perspective must represent their object is a distinctive manner. Let this manner of representing be called the ***second-person mode of presentation***, and let a ***second-person thought*** be a thought that represents its object via a second-person perspective.

Comparing second-person thought to the more familiar examples of self-conscious thought is instructive here. Self-conscious thoughts, like the thoughts that might be expressed by the thinkers by

[23] As some enactivists claim, e.g., Hutto and Myin 2013.

[24] For the purposes of motivating a discussion of second-person thought, all I claim here is that, given common assumptions about the nature of thought and its role in explaining action (and interaction), we are able to occupy second-person stances in virtue of having (at least some) mental states which manifest the second-person perspective. Surely this claim is plausible, if not ultimately correct, thus it seems to be worth developing an account of the second-person perspective in thought.

saying *I am making a mess* or *My pants are on fire*, represent their objects in a distinctive way. Let's call such thoughts first-person thoughts, and the (distinctive) manner in which they represent their objects the first-person mode of presentation. Famously, philosophers contend that some first-person thoughts are irreducibly first-person thoughts, as the cognitive significance of the first-person mode of presentation via which they represent their object cannot be captured by a co-referential third-person mode of presentation. That is, thoughts like *I am making a mess* and *My pants are on fire* seem to differ from thoughts that ascribe the same property the same object, e.g. thoughts like *John Perry is making a mess* or *The man in the mirror's pants are on fire*.[25] Thus, self-conscious thoughts are not merely thoughts about the person who happens to be the subject of the thought. Self-conscious thoughts are irreducibly first-person thoughts that represent the subject of the thought as such.

The same basic template can be applied to second-person thoughts. Second-person thoughts represent their objects via a second-person mode of presentation. An irreducibly second-person thought is one whose second-person mode of presentation cannot be replaced by some combination of first- and third-person modes of presentation without changing the cognitive significance of the thought.[26] But before questions about irreducibly second-person thought can be taken up, the second-person mode of presentation itself must be clarified. For, unlike first-person thought, many philosophers are inclined to think that second-person thought is incoherent, or even if it does exist, it reduces away to some other kind of thought.

If this asymmetry holds, it marks an important difference between the first- and second-person modes of presentation. I don't think that such an asymmetry holds up to scrutiny, but showing why requires first clarifying some features of the second-person mode of presentation. The basic idea is that a second-person mode of presentation is the distinctive mode of presentation through which we represent an agent in the context of a reciprocal interaction (e.g., an interaction in which my action is done in response to yours, yours in response to mind, etc.). To discover what (if anything) is distinctive about this mode of presentation, philosophers have tried to get a handle on the relevant notion of interaction. But, even confining ourselves to accounts of second-person thought, as opposed to the

[25] These examples come from Perry 1979 and Kaplan 1989 respectively.
[26] I am deploying a commonsense notion of "cognitive significance" here. Thoughts differ in cognitive significance when they license different inferences, explain different actions, etc. While I am assuming that thoughts might differ in their cognitive significance despite having the same referent, and labelling such differences as differences in mode of presentation, I am not making assumptions about what such a difference in cognitive significance might amount to.

second-person perspective more generally, there is surprising variation in the accounts that have been produced. For example, when discussing second-person thought, different accounts hold that such thought is fundamentally to be understood in terms of:

> *Communication*: "The phenomenon of the second-person is a linguistic one, bound up with the fact that utterances, as we make them, are typically directed to people, not just made to the cosmos" (Heck 2002, p. 12).
>
> *Common Knowledge & Cooperation*: "Were one person thinks of another that she is F, conceives of her in the 'we minus I' way…., and also the cooperative activity going on is of the face to face character…then the first agent's thought is second person" (Heal 2014, p. 326).
>
> *Mutual Awareness (or Interpersonal Self-Consciousness)*: Second-person thought is "a distinctive way we have of thinking of each other, under a concept that refers *de jure* to its addressee and whose availability depends on standing in a relation of interpersonal self-consciousness with another" (Salje 2017, p. 817).
>
> *Recognition*: Second-person thought involves recognizing others, in the sense that "the relation of free beings to one another is a relation of reciprocal interaction through intelligence and freedom. One cannot recognize the other if both do not mutually recognize each other; and one cannot treat the other as a free being, if both do not mutually treat each other as free" (Fichte 2000, p. 42).

In what follows, I'll argue that each of these accounts captures something important about second-person thought, but that only recognition is capable of furnishing an acceptable general account of the second-person mode of presentation. By considering what goes wrong in each case, I'll develop and defend a recognition-based account of the second-person mode of presentation.

Before doing so, I'll make a brief methodological point. I understand issues of the sort that I am engaged with to consist in a kind of high-level empirical theorizing. Any successful account of second-person thought will ultimately be corroborated, and further developed, by work in the cognitive sciences. That said, second-person thought isn't merely a theoretical term to be postulated and evaluated experimentally. Second-person thought is also a folk psychological kind: if second-person thought exists, and is even moderately interesting, then it will be the sort of thing that we know about already, and is embedded in our practices. Thus, we might reasonably expect people to have intuitions about what counts as a second-person thought, and I intend to try to capture these.[27] While conformity

[27] See Perry 2011 for similar remarks about theorizing the notion of content more generally. I take the discussion of first-person thought to be similar in relevant respects.

with intuition is not a hard and fast criterion, it is generally desirable, and it counts against an account of second-person thought if it seems to misrepresent familiar cases about which we have settled intuitions.

2.2 The Paradigm Interaction Strategy.

Accounts of second-person thought appeal to interaction in order to articulate the distinctively second-person mode of presentation. But what is the relevant type of interaction, and which features of it are important for understanding the second-person mode of presentation? At this juncture, one might pursue two basic strategies. The *Paradigm Interaction Strategy* involves choosing a paradigm kind of interaction, and then using it as a template for understanding the distinctively second-person mode of presentation. Alternatively, the *Functionalist Strategy* involves discovering the cognitive functions that are generally required for interaction (of the relevant sort), discerning which of these functions second-person thoughts might plausibly perform, and using this function to generate an account of the distinctively second-person mode of presentation.[28]

This section will examine two instances of the paradigm interaction strategy. Communication-Based Accounts seek to understand the second-person mode of presentation in terms of paradigmatic communicative interactions, while Cooperation-Based Accounts seek to utilize common knowledge in cooperative interactions to develop an account of the second-person mode of presentation. Each strategy captures some insights concerning the second-person mode of presentation. But, both accounts ultimately fail, as they are unable to provide a general account of second-person thought. I conclude that proponents of second-person thought ought not pursue a paradigm-based account of the second-person mode of presentation. Instead, they are better served by some form of the functionalist strategy discussed in section 3. To see why, first consider the communication-based accounts.

2.2.1 Communication-Based Accounts of the Second-Person Mode of Presentation.

Communication-based accounts begin with the plausible idea that we might articulate an account of the second-person mode of presentation by examining the language used to express second-person

[28] These two strategies probably aren't exclusive. For example, a proponent of the *phenomenal intentionality* thesis might reject both of them, instead claiming that the second-person mode of presentation is constituted by some element of the phenomenal character of interaction experience. In what follows, I'll put the phenomenal intentionality thesis aside, noting that it too might pair nicely with the recognition-based strategy that I endorse in this chapter.

thoughts. For example, if you and I are carrying a sofa up some stairs and you are going too fast, I might have a second-person thought which I express by saying "You need to slow down." The second-person pronoun 'you' displays several interesting features here. It is an indexical, as utterances of 'you' might refer to different individuals depending upon features of the context of utterance. But unlike more permissive indexicals, like 'this' or 'that', 'you' can only be used to refer to agents. Outside of mock uses, which I will put aside for the time being, one cannot felicitously use 'you' to refer to inanimate objects. Most importantly, the word type 'you' has a distinctive character, or standing meaning: utterances of 'you' refer to the addressee of the utterance.[29] Thus, utterances of 'you' are used to address individuals with whom one is thereby speaking.

These features of 'you' are uncontroversial, but don't yet furnish an account of second-person thought. Proponents of *Communication-Based Accounts* are impressed by some further facts about how 'you' functions conversationally, and seek to use such communicative interactions as a paradigm upon which to base an account of the second-person mode of presentation. When a speaker uses 'you' in a conversation, they express a thought about the addressee of the utterance. And when an addressee hears this utterance, she understands the thought in question to be about she herself, which she might express by using the first-person pronoun 'I'. It is this link between 'you' and 'I' that John McDowell seeks to capture in the following:

> Suppose someone says to me, "You have mud on your face." If I am to understand him, I must think and "I"-thought, thinking something to this effect: "I have mud on my face: that is what he is saying" (McDowell, 1984, p. 222).

Here McDowell is making a claim about what a competent speaker of English understands when confronted with an utterance of 'you': their understanding the utterance consists in their coming to have a thought that they might express by using 'I'. But by making further assumptions, one reaches conclusions about the thought expressed by utterances of 'you', and thus the second-person mode of presentation. Communication-based accounts use these tools—the semantic features of the second-person pronoun 'you' and its role in successful communication—to construct an account of the second-person mode of presentation.

[29] Here, and in the remainder of the paper, I will assume that we are focused only upon the singular use of 'you'. There is also a plural use of 'you', though these two words are plausibly taken to be homophones.

All versions of the communication-based account model the second-person mode of presentation closely upon the meaning of the second-person pronoun 'you'. But different philosophers have taken communication-based accounts to support quite different conclusions about second-person thought. For example, Richard Heck articulates a standard view of the meaning of 'you' when he says that "as a matter of its standing meaning, an utterance of 'you' refers to the person addressed in that utterance" (Heck 2002, p.12). Enforcing a strict analogy between the meaning of second-person language and the intentionality of second-person thought, yields the following account of the second-person mode of presentation: much like 'you', the second-person mode of presentation represents its object as the person addressed by the thought. Further, address has a particular meaning here. When an agent addresses another agent, she performs an intentional action which isn't merely about the other agent, but rather is directed at the other agent. This generates a problem, as thoughts don't seem to address their objects in this sense. More pointedly, thoughts don't even seem to be the right sort of thing to address in the first place. As such, the second-person mode of presentation might either address its objects, or represent its object as being addressed by thought. Neither of these is plausible, thus there are no second-person thoughts.

Some proponents of communication-based accounts challenge this conclusion. They grant that although in some sense there are no irreducibly second-person thoughts on this account, it needn't mean that all thoughts aptly expressed using 'you' are merely third-personal. Guy Longworth and Sebastian Rodl each accept that there are no irreducibly second-person thoughts, but argue that some second-person thoughts are nonetheless distinctive from normal third-person thoughts, as some second-person thoughts are also first-person thoughts, albeit first-person thoughts not directed at their thinker. Reflecting upon successful communication using utterances of 'you' and 'I' clarifies the argument for these "other-directed first-person thoughts" (Longworth 2014, p. 289). When I express a thought by using the second-person pronoun 'you', and you understand it, what you understand is something that you would express by using the first-person pronoun 'I'. But this use of 'I' is not accidental; your understanding consists in having an essentially first-person-indexical thought. Further, they assume a *naïve conception of communication* on which successful communication consists in a speaker uttering U to express their thought T, and the hearer coming to believe T on the basis of hearing and understanding U. Combined, this yields the conclusion that the thought I expressed with my utterance of 'you' is an essentially first-person indexical thought. But the thought is not about me, it is a thought

about you. Thus, a thinker can have, and communicate, essentially first-person indexical thoughts that are not about the thinker herself.

For all of their differences, each of these versions of the communication-based account suffers from the same problem: the paradigmatic interaction they've used to model the second-person mode of presentation upon is ill-chosen. Communication-based accounts model the second-person mode of presentation upon the character of 'you'. The resulting accounts of the second-person mode of presentation are modeled upon communicative interactions that deploy 'you', and in particular, such accounts take the notion of linguistic address from the character of 'you' and builds it in to the second-person mode of presentation.

Unsurprisingly, thoughts don't address their objects, or represent their objects as being addressed, thus there are no irreducibly second-person thoughts. In itself, this conclusion isn't problematic. The problem is that the conclusion is reached by employing an unduly weak account of the second-person mode of presentation. On the communication-based account, second-person thought embodies something akin to a category mistake. Linguistic address is something that agents do.[30] For example, one agent might address another by making a certain kind of utterance. Sub-agential states don't address agents in this manner. It isn't even clear what that would mean: to the extent that the state itself could address an agent, one feels pushed towards the conclusion that the state is itself an agent. But it isn't. This conclusion ought to leave us wondering if better accounts of the second-person mode of presentation are available.

If there are, then communication-based accounts have turned out to be wrong, but in an uncharitable way. And in doing so, they leave interesting options for how to understand the second-person mode of presentation unexplored. My suspicion is that communication-based accounts offer a natural treatment of the second-person stance, but the accounts that I've discussed seek to articulate an account of the second-person mode of presentation deployed by second person thought. Thus, at heart, these accounts take linguistic address—something that agents can do—and attempt to use it to model the second-person mode of presentation. But modes of presentation are potentially manifested by states of an agent (thoughts), but not the agent herself. Accordingly, communicative interactions,

[30] Salje 2017 argues that second-person skeptics err in presuming that all address is linguistic address. She argues that some address is non-linguistic, and that thoughts can address in a non-linguistic sense. I discuss Salje's position in section 3.

including the notion of linguistic address, are not fit candidates for articulating an account of the second-person mode of presentation. Second-person thoughts don't address, or represent their object as being addressed, but there was never reason to think that they even *could* have. Unless we are forced to endorse a communication-based account,[31] or simply run out of other options, we should look elsewhere to understand the second-person mode of presentation.

2.2.2 Cooperation-Based Accounts of the Second-Person Mode of Presentation.

Maybe the communication-based account simply chose the wrong type of interaction to use as a paradigm. In place of communicative interactions, *Cooperation-Based accounts* use cooperative interactions to build an account of the second-person mode of presentation. Highly cooperative animals, like humans, have distinctive capacities that allow us to engaged in cooperative actions. We can form collective intentions that represent not only what you will do in a cooperative interaction, and what I will do, but what we will do. Cooperation-based accounts attempt to utilize this cognitive machinery in understanding the second-person mode of presentation. In particular, these accounts contend that we can understand the distinctively second-person mode of presentation as derivative upon the kinds of thoughts that we have during face-to-face cooperative actions. In Jane Heal's slogan: "'you' is 'we minus I'" (Heal 2014, p. 317).

To illustrate the cooperation-based account, let's further unpack Heal's account. Many philosophers contend that first-person thought provides reasons for the subject to act. For example, suppose that I know that anyone who is F should perform action A. I might learn any number of things about myself (e.g. that the tallest philosopher in the grocery store is F), but it isn't until I learn that "I am F" (an irreducibly first-person thought) that I gain a reason to perform A.

Similarly, in co-operative action, there is a kind of first-person plural intending and judging. For current purposes, let cooperation be "a kind of action or activity where agents…have a common purpose and hence common knowledge of their circumstance and of what they are doing" (Heal 2014, p. 322). Assuming that two agents are cooperating,[32] these agents "have a common purpose, and so know themselves to be acting as a pair" (*ibid.*, p. 323). When acting together, two agents form joint

[31] For example, should one's theory of content imply that thought and language are connected in the manner that communication-based account presuppose, then one might be *ipso facto* committed to such an account.
[32] Cooperation could involve more than 2 agents, but, following Heal, I will simplify the discussion by limited the cooperative interaction to 2 agents.

intentions, intentions about what 'we' will do (as they might put it), which play "the distinctive role of giving reason for contribution to the cooperative activity" (Heal 2014, p. 317).

In cooperative action, the agent's common knowledge plays a role in rationally explaining the ensuing action. Heal gives the following example (*ibid.*, pp. 323-324) of cooperative reason that would rationally explain a joint action. Imagine two agents, Sam and Sara, who have the following common knowledge:

(4) Any co-operating pair which is F should V1.
(5) If either of a pair is G then the pair is F.
(6) For a pair to V1 both members must V2.

Next, imagine that Sam becomes G, and this too is common knowledge. From this it should follow that Sarah should V2. But note that Sarah might learn any number of things about the group comprised of her and Sam (e.g. that the tallest philosopher in the grocery store and the shortest philosopher in the grocery store are F). It isn't until she learns that "We are F," a thought which deploys a distinctively first-person plural mode of presentation, that she gains the right kind of reason to perform V1. Thus, if what Sarah learns when Sam becomes G is to be capable of explaining her action, what she learns is (from Sarah's point of view):

(7) [We minus I] is G.

For the moment, treat '[We minus I]' as a place holder. Assuming that each party performs the relevant inferences, it follows that each of the following are now common knowledge too:

(8) We are F
(9) We should V1

And from this, Sarah can conclude that:

(10) I'll V2!

Intuitively, Sarah's thought in (7) is aptly described as a second-person thought. It is a thought about an agent, Sam, with whom she is performing a cooperative interaction. Further, she might naturally express this thought by saying "You are G!" If so, Heal contends, this example clarifies the nature of the second-person mode of presentation: this mode of presentation represents another as a partner in cooperative action, a "we minus I." Thus, Heal's cooperation-based account effectively pulls the

second-person mode of presentation out of something that both agents know, the common-knowledge structure that drives cooperative action.

The resulting account is appealing, but faces a key problem: not all interactions in which we plausibly have second-person thoughts are cooperative ones. Although many of the most intuitively plausible cases of second-person thought involve agents cooperating to perform some action, many second-person thoughts occur during antagonistic interactions between agents. Imagine two boxers carefully circling each other in a ring, or a prey animal trying to remain still so as to avoid being seen by a predator. In such cases, there is no common purpose between the two agents, hence there is no common knowledge of "what they are doing." In these cases, there is a perfectly natural sense in which "they" aren't doing anything at all, and the outcome of a boxing match, or an interaction between predatory and prey, isn't well explained as a piece of joint reasoning towards some outcome.[33] This isn't to say that what each boxer does, or what the prey animal does cannot be rationally explained. So long as each animal (boxer and prey) can act for reasons, it seems like they can act for reasons in antagonistic interactions. More importantly, there is no reason to think that these boxers or animals cannot have second-person thoughts about their interlocutor. Surely they can! But these second-person thoughts cannot be understood in Heal's framework, thus the framework seems to be is too restrictive to offer a general account of second-person mode of presentation.

Heal might reply by claiming that she doesn't mean for her account to capture all cases thought that manifests the second-person perspective. She says of her account of second-person thought that "I do not claim to show that the thoughts so identified are the only things which could be called 'second person thought'. Perhaps, there are yet other kinds of thought which also deserve the label" (Heal 2014, p. 318). However, this reply comes at a serious cost. While Heal might have identified a cooperative second-person mode of presentation, she hasn't characterized the second-person mode of presentation in its most general form. And it is precisely this most general characterization that we want to use when evaluating arguments for or against the irreducibility of second-person thought.

[33] One might suspect that there is some loosely defined activity that both parties are engaged in. There is might, however the point of the objection is that if the activity lacks a common-knowledge structure, then it can't be used to furnish the second-person mode of presentation.

Each version of the Paradigm Interaction Strategy described in this section faces problems, and should be rejected as a general account of the second-person mode of presentation. In the case of communicative interactions, accounting for second-person thoughts during non-linguistic interactions (which don't involve linguistic address) poses problems. In the case of cooperative actions, problems arise in accounting for second-person thoughts during non-cooperative actions (which don't involve common knowledge and shared purpose). The problems facing each account look to be traceable to the paradigm-based strategy for articulating the second-person mode of presentation.

But my intended conclusion of in this section is not merely that we should choose a different paradigm. Instead, I suspect that any version of the paradigm interaction strategy will either (a) mistakenly impute features of agents onto thought, or (b) offer too narrow an account of the second-person perspective. This results from attempting to articulate an account of the distinctively second-person mode of presentation in terms of a paradigm form of interaction, despite the fact that the explanatorily appealing features of these interactions are either (a) most naturally understood as features of agents (not their states), or (b) not shared by all interaction in which one might have second-person thoughts. The way forward is to put paradigms aside, and pursue a different strategy for articulating the second-person mode of presentation.

2.3 The Functionalist Strategy.

Thinking of the second-person perspective in functional terms offers a promising alternative. Instead of constructing a blueprint for the second-person mode of presentation based upon a paradigm type of interaction, one instead attempts to clarify the function that second-person thought plays in interaction, and to use this function in generating an account of the second-person mode of presentation. If done well, the functionalist strategy won't confuse agent-level features for features of thought because the function is to be specified in sub-agential terms. Further, if a suitable function can be found, it avoids overgeneralizing from features found only in a subset of interactions. Thus, the functionalist strategy is poised to avoid the vices of paradigm interaction accounts.

If second-person thought has no discernable function in interaction, then the strategy won't be fruitful. But things don't look quite so dire. Upon reflection, the philosophical literature on second-person thought seems to be circling a small set of functions that it might perform, albeit while often lapsing into precisely the sorts of accounts that section 2 contends should be avoided. In her useful

discussion of second-person thought, Naomi Eilan talks about second-person relations, which are "instances of the relation in which two people stand to each other, in virtue of which use of the second person is possible and appropriate" (Eilan 2014, p. 266). The basic idea is that during interactions of the relevant type, which one might variously call "engaged" or "reciprocal" interactions, participants to the interaction stand in a certain kind of cognitively-mediated relation to each other. Should we avoid the temptation to lapse into characterizing second person relations in terms of apt use of 'you' or in terms of cooperative action (e.g., by talking about shared action, acting together, etc.), it appears that many of these philosophical attempts to understand second-person thought have some core common understanding of the function played by second-person thought. The functional strategy attempts to clarify this core in order to better understand the function of second-person thought.

In which terms should we specify this function? I'll consider two related, but different, approaches to understanding the function of second-person thought: mutual awareness and recognition. Because the recognition-based account avoids a key problem that faces the awareness-based account, and connects naturally to work on the second-person stance, I conclude is preferable as account of the second-person mode of presentation. First, however, we will examine accounts of the function of second-person thought based in mutual awareness.

2.3.1 Mutual Awareness Account.

The mutual awareness account contends that second-person thought represents its subject and its object as being mutually aware of one another. The function of the second-person mode of presentation is to be captured by spelling out this kind of mutual awareness. More specifically, the claim is that second-person thought involves a kind of *interpersonal self-consciousness*, in which the subject of the thought is aware that they feature in another person's consciousness. To clarify this account, I first unpack the account of mutual awareness that is being employed, and then introduce Salje 2017's account of second-person thought.

Interpersonal self-consciousness involves a distinctive kind of self-awareness, namely the "awareness that you are being represented by another persona as a self-representing subject" (Peacocke 2014, p. 239). This phenomenon is quite routine, but for a salient (and particularly awkward) example, suppose that you and I are sitting in a restaurant booth gossiping loudly about someone. As we gossip away, I notice that there are other people around us, but they seem not to be paying any attention. However,

mid-way through the conversation, I notice that the object of our gossip has been sitting in the booth right next to us! As they get up to leave, and shoot me a scathing look, I experience a state of interpersonal self-consciousness: I am aware that I feature in their thought, not merely as an object, but as me myself.

Regardless of its potential application to second-person thought, there seems to be ample reason to accept interpersonal self-consciousness as a variety of self-conscious thought. Having noted this, one might then put the notion to work in generating accounts of phenomena like joint action, linguistic communication and second-person thought.[34] Peacocke 2014 and Salje 2017 both motivate the mutual awareness account of second-person thought by considering the apt conditions for expressing a thought by using 'you'.[35] Salje notes that there are distinctive patterns in our apt use of 'you' to express thoughts, as opposed to other co-referential terms: "there is a reason that I use 'you' to refer to you, when I do. My use of the second-person pronoun signals a special set of awareness relations holding between us" (Salje 2017, p. 821). After unpacking this, and arguing that there are four conditions on the apt use of 'you', Salje concludes that "in sincere and successful uses of the second-person pronoun, I am aware that I feature, myself, in your consciousness as a conscious subject, because I am aware that you are aware of me being aware of you. My use of the second-person pronoun signals my interpersonal self-consciousness with respect to you" (Salje 2017, p. 826).

Putting aside claims about the apt use of 'you' to express second-person thoughts, the resulting account of the second-person mode of presentation looks something like the following:

> X has a second-person thought S about Y iff S represents X's interpersonal self-consciousness with respect to Y.
>
> X's interpersonal self-consciousness with respect to Y consists in:
> (i) X is aware of Y.
> (ii) Y is aware of X.
> (iii) X is aware that Y is aware of X.
> (iv) Y is aware that X is aware of Y.

[34] As Peacocke 2014 does.

[35] This is problematic for the reasons introduced in the previous section. That said, I will put aside my objections to communication-based accounts for the moment. The mutual awareness account that Peacocke and Salje ultimately deliver can be divorced from their particular (communication-based) means of motivating it, and I will criticize it independently.

Thus, the second-person mode of presentation consists in a kind of mutual awareness, which is further spelled out in terms of conditions (i)-(iv). It is worth pointing out two features of this account. First, there are two broadly different ways of developing the mutual awareness account of second-person thought. One might argue that these conditions of mutual awareness are part and parcel of the content itself. Alternatively, philosophers of a more externalist bent might argue that (i)-(iv) are enabling conditions for second-person thought. The permissive understanding of mode of presentation that I am employing here is compatible with either reading, so I won't take sides here.

Second, the mutual awareness account doesn't by itself take sides on issues of the irreducibility of second-person thought. Proponents of the mutual awareness account disagree about the irreducibility of second-person thought, with Peacocke 2014 offering a reductive of account of second-person thought, and Salje 2017 arguing for a form of irreducibility. Again, this is an appealing feature of the account, as it seems like an account of the second-person mode of presentation ought not *by itself* settle issues about the irreducibility of second-person thought.

The mutual awareness account has much to recommend it, but I contend that it ultimately errs in focusing exclusively on awareness. It is doubtlessly the case that awareness and interaction are intimately related. But focusing only on awareness yields an account that is inadequate. Only a very thin sense of awareness is plausibly necessary for second-person thought, and mere mutual awareness doesn't seem to be sufficient for second-person thought. I'll argue for each claim in term.

First, it seems like one can have second-person thoughts about other agents, even if those agents are unconscious. Imagine that you've built a robot that exhibits complex, adaptive behavior and is a pretty good chess player. But it isn't conscious, and, having built it, you are quite aware of this. You are playing a game of chess with this robot, and it falls for your carefully set trap. Thrilled, but fully aware that the robot still lacks the capacity for conscious awareness, you have a second-person thought about it—*you fell for it*.

One could reply that despite your knowledge of the robot's lack of consciousness, in the heat of the moment you temporarily slip into thinking of it as a conscious agent. But I suspect that proponents of the mutual awareness account would instead, simply grant my point and offer a thinned-out notion of awareness. While there is a tendency to smuggle various capacities into the term 'awareness', there

needn't be. On this account, mutual awareness doesn't mean conscious awareness, or attention, it simply means having thoughts that represent their object in a certain sort of way. This is a deflationary way of reading the claim that "I am aware that I feature, myself, in your consciousness as a conscious subject, because I am aware that you are aware of me being aware of you" (Salje 2017, p. 826). Of course, there is a lingering worry that the awareness account, so understood, is too thin—is second-person thought really just subjects of thought being mutually aware in this thin sense? If so, it remains unclear how to connect this sort of representational phenomenon to anything of further interest.

A more pressing problem is that one can stand in this thin sense of mutual awareness to another subject without having a second-person thought about them. If so, then something about second-person thought is missing from the mutual awareness account. To see this, imagine a creature with the capacity for thought, but absolutely no capacity for movement, action or expression. It doesn't orient itself, talk, move, or even have a face with which to express itself. You can think about this creature, and it can think about you, but you can't really interact with it. Moreover, this is all salient to you as you form a thought about this creature that meets conditions (i)-(iv) above.

Is the resulting thought about the creature a second-person thought? It seems to me like it isn't.[36] Although the creature might think of you, and you of it, you don't think about it in the context of an interaction with the creature. Thus, your thought about the creature fits the conditions for interpersonal self-consciousness, but it isn't a second-person thought. Something seems to be missing. Particularly, it seems like the intuitive idea that second-person thought is tightly linked to interaction has dropped off. Thus, while interpersonal self-consciousness might well be linked to second-person thought, and self-consciousness might still play a distinctive role in second-person thought, second-person thought involves more than mere interpersonal self-consciousness.

None of this suggests that interpersonal self-consciousness is not a real phenomenon, nor that it can't be put to explanatory work. Rather, it suggests that the functional role of the second-person mode of presentation cannot be captured purely in terms of interpersonal self-consciousness. Any form of mutual awareness that second-person thought plausibly involves is too thin. Further, second-person

[36] If necessary, we can imagine that this creature has rudimentary capacities for perception, and that it forms its thoughts about you on the basis of perceiving you. But it still can't act: it doesn't have a face, it can't orient towards you, etc. I don't think that granting rudimentary perceptual capacities *alone* makes much difference to the case.

thought seem to be connected to interaction in a way that is not captured by the mutual awareness account. The basic point is that although awareness and interaction are doubtlessly related, the function of second person thought seems to involve more than merely representing mutual awareness. Thus, the awareness-based account of the second-person mode of presentation cannot succeed.

2.3.2 Recognition-Based Account.

The objections to the awareness-based account are suggestive, as it seems that second-person thought is connected to representing the agency of another in the context of an interaction. This is unsurprising, given the intuitive starting point of theorizing about the second-person perspective generally: the second-person perspective is the perspective of a participant in an interaction, as opposed to a mere observer. This "participant perspective" doesn't seem to have anything specifically to do with language, cooperation, or mutual awareness, but with *participation* in an *interaction*. It is by returning to this intuitive starting point, and further clarifying both the kind of interaction in question, and the function that second-person thought performs in such interactions, that we might better characterize the second-person mode of presentation.

To begin, we must clarify the type of interaction in question. Not just any form of interaction is apt or the purposes of capturing the "participant perspective" of second-person thought. For this purpose, we must focus upon the kinds of interactions which take place between interlocutors. There is a very loose sense of interaction, in which I can interact with the inanimate world around me. For example, I interact with a cup when I pick it up. Similarly, I might interact with another agent who isn't paying attention to me, or simply takes me to be a part of the inanimate environment. None of these cases are interactions in the relevant sense. It isn't that the thinker isn't participating, but that they don't really have an interlocutor with who they are engaged in a kind of reciprocal interaction, in which I act knowing that you will respond to my action, you act knowing that I will respond to your action, etc. It is in this sense that interaction requires interlocutors, or at least taking someone else to be an interlocutor.

Whatever else might be distinctive about these kinds of reciprocal interactions, it seems like they involve a characteristic way of thinking about one's interlocutor(s). In such cases, second-person thought represents our interlocutors not as objects, nor merely as agents, but as agents with whom we are interacting. To illustrate this, contrast two experiences: the experience of interacting with other

agents and the experience of merely observing them. Imagine walking down a narrow sidewalk as someone walks towards you. Looking directly at each other, each of you moves to your right to get out of the way of the other person. In this case, as you walk past the other agent, you have a thought about them that lead to your moving to the right. It is plausibly a second-person thought.

Contrast this experience with one in which you believe yourself to be a ghost—invisible and immaterial—but, a ghost that is fairly habituated to the kind of interactions that material beings have. You still take the stairs, use doors, and avoid floating through people. Thus, your "ghostly" interactions otherwise proceed in exactly the same manner as your embodied ones might have. As ghostly-you walks down the hall, the agent seems to look directly at you, and seems to move to his right as if he were getting out of your way. And of course, you do look at him, and you walk to your right as if getting out of his way. But you don't believe that the other person can see you. And you, being immaterial, could walk right through them anyway. Your thought in the second case seems to be different from your thought in the first case, but the relevant difference cannot be explained in terms of which thoughts you have, nor in terms of what your thought is about. By stipulation, these are the same in each case. Rather, it seems that what changes between the two cases is how you think about the other agent. When you take yourself to be engaging in a coordinated interaction with them, you think of them in a particular way. And when you are merely observing them acting, even acting in the exact same way they would have if you were interacting, you think of them in a different way.

Of course, the connections between thought and experience are complex, and this contrast argument only establishes the plausibility of a distinctive kind of thought about other agents. This kind of thought deploys a distinctive kind of representational capacity which can be understood as a claim about the function of second-person thought. I'll say that second-person thought involves *recognizing* an agent as a partner in interaction. The *Recognition-Based Account* of second-person thought contends that this recognition function of second-person thought can be used to articulate an account of the second-person mode of presentation.

At this point, one might worry that the recognition-based account is duplicating the error made by communication-based accounts. The most familiar philosophical uses of the term 'recognition' describe a stance that agents take up in relation to other agents. Political philosophers sometimes speak of recognition as a kind of respect for other's right to self-determination (e.g. Taylor 1994),

while moral philosophers sometimes speak of recognition (or recognition respect) as a way of registering the dignity or authority of other people (e.g., Darwall 2006). Whether plausible or not as claims about moral and political philosophy, these other uses of the term recognition are claims about stances taken up by agents, and might be problematic as claims about the function of second-person thought. But neither of these capture what I am claiming is the function of second-person thought.

As a point of clarification, therefore, it is worth contrasting my use of recognition *as a function of second person thought* with some other similar uses in the literature. The uses of the term recognition in moral/political philosophy might seem to capture well the reciprocal nature of the kinds of interactions that I'm discussing. For example, recall Fichte's claim that:

> "the relation of free beings to one another is a relation of reciprocal interaction through intelligence and freedom. One cannot recognize the other if both do not mutually recognize each other; and one cannot treat the other as a free being, if both do not mutually treat each other as free" (Fichte 2000, p. 42)

This is apt to sound a lot like the kinds of interactions that I'm discussing, and the relevant way of thinking about another agent. Indeed, such discussions of recognition serve as the inspiration for a recognition-based account of the second-person mode of presentation. But, as I use the term, 'recognition' describes the role that second-person thought plays in interaction. A thought recognizes an agent when it represents the agent in a particular way, viz. as an agent with whom I am interacting. Thus, second-person thoughts recognize interlocutors in an interaction because the second-person mode of presentation that such thoughts deploy represents the agents in this manner. On the recognition-based account, the second-person mode of presentation represents its object as *an agent with whom I am interacting*.

So, the operative sense of recognition is related, but not identical, to the one drawn from moral/political philosophy. Neither is it the notion of recognition that it is often used in the cognitive science literatures that study object recognition, or facial recognition. My use of 'recognition' isn't necessarily more normatively-laden than typical uses in the visual cognition literature, rather it names a different representational capacity. The recognition function of second-person thought does not have to do with representing the familiarity of some object, but rather with how it represents the thing—as a rational agent, or locus of agency, that is interacting with me myself.

The second-person mode of presentation has two further features of interest. First, it involves representing its object as an agent. In this way, it is reminiscent of the meaning of the second-person pronoun 'you'. But I should be clear about what I mean here. The second-person mode of presentation doesn't merely represent something as falling under a kind AGENT, plus some stuff about interaction. Rather, the second-person mode of presentation represents another object as a rational agent, or a locus of agency that acts for reasons in much the same manner that I do. As such, one might say that the second-person mode of presentation represents its object agent as "another I."

Second, the second-person mode of presentation represents its object as an agent with whom *I* am interacting. Thus, a tight connection between the first- and second-person perspectives is preserved. Again, this is reminiscent of the close connection between 'you' and 'I' when communicating linguistically. Both of these features parallel those articulated by communication-based accounts. But on the recognition-based account, these features are not reached by appealing to the language used to express these thoughts. Instead, they are reached by considering the function played by the second-person thoughts themselves. As such, the recognition-based account is not at risk of illicitly imputing agent-level features of communication into an account of the second-person mode of presentation. Instead, it seeks to merge the insights gleaned from previous accounts, all the while eliminating some of their objectionable features.

All told, the recognition-based account of the second-person mode of presentation is rooted in two claims: (1) that a certain type of reciprocal interaction is relevant to understanding the "participant perspective" of second-person phenomena, and (2) that the characteristic function that second-person thoughts play in such interactions is recognition. By appeal to this recognition function, the recognition-based account claims that the second-person mode of presentation consists in representing another agent as a rational agent with whom I am interacting. This is neither to appeal to the sense of "recognition" familiar from political philosophy, nor the sense familiar from work on visual cognition. Instead, recognition is a distinctive kind of representational capacity, which enables us to have reciprocal interactions of the sort that consist in taking up the participant stance.

2.4 Conclusion.

The second-person perspective comprises a cluster of philosophically interesting phenomena. By focusing upon second-person thought, or thought that represents its object by employing a second-

person mode of presentation, this paper evaluates some main accounts of the second-person mode of presentation. It reaches two main conclusions. The first conclusion is methodological in nature, and concerns how philosophers ought to study the second-person mode of presentation. Philosophers studying the second-person mode of presentation shouldn't proceed by using paradigm instances of interaction as blueprints for the second-person mode of presentation, as this produces overly narrow accounts. Instead, philosophers are better served by understanding the second-person mode of presentation in terms of the function that second-person thought performs in interactions.

The second conclusion concerns how to understand this function. While there are various ways of construing the function performed by second-person thought, the most general function that it performs involves recognizing another agent in the context of an interaction. Using this function to develop an account of the second-person mode of presentation, I've argued for a *recognition-based* account, on which the second-person mode of presentation represents its object as an agent with whom I am interacting.

It is important to appreciate both what these conclusions accomplish, as well as their limitations. Ultimately, this paper aims to clear the ground for a study of second-person thought. There are philosophically interesting connections between the second-person perspective, language and normativity. But, I've argued, it is a mistake to try and understand second-person thought exclusively in terms of such connections. Doing so either collapses the distinction between the second-person standpoint and the second-person thought, or it yields an implausibly restricted account of the second-person mode of presentation. The functionalist strategy that I've defended shows how one might accept claims about the connections between the second-person perspective, normativity and language, while nonetheless crafting a plausible account of the second-person mode of presentation.

But I haven't yet given a full account of second-person thought. What I've delivered is best understood as a functional *gloss* on the second-person mode of presentation. After all, I haven't yet said what a mode of presentation is, and different accounts would surely yield different accounts of the second-person mode of presentation. Should one wish to fill out the functional story I've given in more detail,

more needs to be said about what modes of presentation are, including how they are to be individuated.[37]

But, although not fully specified, the recognition-based account already has philosophically interesting features. For one thing, it offers an account of second-person thought that navigates between elimination and over-intellectualization. Due to the tight relation between the second-person standpoint and interaction, it is tempting to collapse second-person thought into something psychological but non-representational (e.g. affect and emotions are sometimes alleged to be like this), or into a feature of the interaction itself (e.g., the dynamics of meaningful, face-to-face interaction). Alternatively, it is tempting to over-intellectualize the second-person perspective, and understand it exclusively in terms of the second-person stance. Both temptations ought to be resisted, and second-person thought offers a means of doing so. The account that I've developed here avoids each of these temptations, and points towards potentially fruitful questions about how the second-person mode of presentation is realized by thought, and how recognition is accomplished by cognitive systems.

More generally, crafting an adequate account of second-person thought is the first step towards evaluating arguments for the claim that thought does, or doesn't, manifest an irreducibly second-person perspective. I have given no such argument here, which is ultimately a good thing—a satisfactory account of second-person thought shouldn't *by itself* settle the issue of the irreducibility of second-person thought. A separate argument, or set of arguments, ought to be required to satisfactorily answer questions about the irreducibility of second-person thought. However, carefully evaluating accounts of the second-person mode of presentation keeps us from smuggling unnecessary, and potentially problematic, commitments into an account of second-person thought.

[37] Chapter 5 gestures at one way of more fully spelling out this functional story.

Chapter 3: Irreducibly Second-Person Thought

Imagine sitting down for lunch at a busy airport. As you raise a sandwich to your mouth, someone nearby issues a stern command: "Stop that this instant!" Perplexed, you look up to see a woman glaring at you, and come to believe that she told you to stop eating. In the context of your ongoing interaction with this woman, you might naturally express your dawning realization by addressing her, "You told me to stop eating." But regardless of how you might express yourself, your thought about the woman seems to represent her in a distinctive manner. Given that you take yourself to be party to an ongoing interaction with this woman, and that this is something about which you could be mistaken, I contend that your thought recognizes her not merely as an object, nor merely as an agent. Rather your though represents the woman as *an agent with whom I am interacting*. As such, the thought is naturally understood as a second-person thought.

To further illustrate, suppose the scenario continues as follows. After forming this thought about the woman, you realize that she isn't looking at you, but just past you. Following her gaze, you turn around and see a sulking child. You now realize that the woman's command wasn't directed at you, but rather that you were caught in between a misbehaving child and their admonishing mother. Appreciating this, your earlier thought now seems to misrepresent the scenario in (at least) two ways. The woman's speech act wasn't directed at you, hence she never told *you* anything at all. Further, you initially experienced the woman as a partner in interaction, and, I contend, your initial thought represented her as an agent with whom you were engaged in an interaction. But you weren't so engaged. Thus, your thought was also inaccurate in how it represented the woman: you mistook her as the other half of a reciprocally interactive exchange, but she was never engaged in this kind of interaction with you.

This case, which I'll call *Airport*, involves a second-person thought, e.g., a thought about (as the subject might put it) an agent with whom I am interacting. Little would hinge on recognizing second-person thought if the second-person mode of presentation it deploys can be reduced away. The second-person mode of presentation reduces to the first- and third-person modes of presentation if the cognitive significance of the second-person mode of presentation can be captured without remainder by some combination of first- and third-person modes of presentation. If one can show that no such

reduction can be made to work, then one shows the irreducibility of the second-person mode of presentation. I'll call a thought that deploys an irreducibly second-person mode of presentation an *irreducibly second-person thought*.

This essay argues for the irreducibility of second-person thought. The argument is simple: *mutatis mutandis* the reasons that favor accepting irreducibly first-person thought also favor accepting irreducibly second-person thought. Well-known arguments of the sort given by Hector-Neri Castaneda, John Perry and David Lewis,[38] contend that irreducibly first-person thought (variously called essentially indexical thought, self-locating thought, or *de se* thought) plays an indispensable role in explaining action and underpins a distinct kind of self-knowledge. Likewise, I argue that irreducibly second-person thought plays an indispensable role in explaining interaction with other agents and underpins a distinct kind of knowledge of these other agents.

But while irreducibly first-person thought is now widely recognized, irreducibly second-person thought enjoys no such popularity. There are two different reasons for this: while some philosophers accept irreducibly first-person thought, and reject only the second-person variant, others reject both irreducibly first- and second-person thought. In response to critics of both kinds, I contend that the arguments for irreducibly first- and second-person thought stand and fall together. My aim here is to defend a conditional claim: if the irreducibly of first-person thought is shown by the kinds of arguments offered by Castaneda, Perry and Lewis, then the irreducibly of second-person thought is shown by the analogous arguments I offer.

The remainder of the paper has the following structure. *Section 1* introduces second-person thought in more detail, distinguishing two ways of understanding the second-person mode of presentation, and develops an account of irreducibly second-person thought. *Section 2* articulates an argument for irreducibly first-person thought drawn from John Perry's *Sloppy Shopper*, and develops an analogous argument for irreducibly second-person thought. *Section 3* offers a parallel treatment of Lewis' example of the *Two Gods*. These arguments support the conditional thesis that if such arguments establish the irreducibility of first-person thought, then the analogous arguments I offer establish the irreducibility of second-person thought. *Section 4* considers two main attempts to reduce second-person thought

[38] E.g., Castaneda 1966 and 1967, Perry 1977 and 1979, and Lewis 1979.

without rejecting the commitment to irreducibly first person thought. *Section 5* concludes the paper by returning to skepticism about irreducibly first- and second-person thought, and arguing that the conditional claim is interesting, regardless of whether or not one accepts its antecedent.

3.1 A Generic Account of Second-Person Thought.

Before arguing for the irreducibility of second-person thought, this section offers a generic characterization of second-person thought, and illustrates two different ways of understanding what the irreducibility of second-person thought might consist in. But before doing so, I'll note the importance of crafting a generic account of second-person thought before arguing for its irreducibility. An adequate characterization of second-person thought should be *generic*, in the sense that it meets the following two desiderata. First, it shouldn't rule out the existence of irreducibly second-person thought by definition. Whether or not second-person thought reduces to some combination of first- and third-person thought should be a substantive finding, and merely not an immediate implication of stipulated definition. Second, it should accrue minimal commitments to theoretical entities that are contested in this context, like Fregean senses or structured propositions. While an account of second-person thought ought to leave open the option to understand thought in terms of such entities, it shouldn't mandate it. To the extent that it can be avoided, irreducibly second-person thought ought not be rejected merely because one rejects some theoretical entity in terms of which it is understood.

To meet these desiderata, I'll offer a generic way of understanding the irreducibility of personal perspectives in thought. Because discussions of irreducibly first-person thought are more familiar in the philosophical literature, I'll begin by introducing the terminology as it applies to such thought. The paradigmatic cases of irreducibly first-person thoughts are *self-locating* thoughts about who, where and when one is. Familiar examples include John Perry realizing that he himself is making a mess, David Kaplan realizing that his own pants are on fire, and Ernst Mach realizing that he sees himself in the bus mirror.[39] Each of these thoughts is a *first-person thought*, meaning not merely that it refers to the thinker of the thought, but that it represents this object in a distinctive manner. Specifically, I'll say that a first-person thought represents its object via a *first-person mode of presentation*. Via this mode of presentation, the subject "is aware of themselves as *themselves*; it is manifest to them that they themselves are the object of awareness" (Smith 2017). Philosophers disagree about exactly how to

[39] Perry 1979, Kaplan 1989 and Mach 1959 respectively.

understand the nature of modes of presentation,[40] but this doesn't matter for current purposes. What is philosophically interesting about the first-person mode of presentation is that its "cognitive significance" is allegedly not capable of being captured by any co-referential third-person mode(s) of presentation. If so, I'll call it an *irreducibly first-person mode of presentation*. Further, let an *irreducibly first-person thought* be a first-person thought that represents its object via an irreducibly first-person mode of presentation.

Sections 2 and 3 clarify and expand upon two classic arguments for irreducibly first-person thought, which is variously called self-locating thought, essentially indexical thought and *de se* thought.[41] My framing of these arguments seeks to avoid the use of these more specific terms, as they often carry with them commitments, or seeming commitments, that pull against the desiderata of genericity. For example, John Perry's use of 'essentially indexical thought' builds in to it a (seeming) commitment to a link between the nature of the thought being expressed and the language used to aptly express it. Alternatively, Lewis's use of '*de se* thought' carries with it a commitment to Lewis's way of understanding the truth-conditions for such thoughts in terms of self-ascribed properties. By framing these arguments in terms of modes of presentation, I mean to sidestep the issues that arise for these more familiar, but also more specific, accounts of irreducibly first-person thought.

Nor do I think this is fundamentally at odds with philosophical orthodoxy. In the afterward to his paper "Frege on Demonstratives," John Perry notes that:

> "In claiming that demonstratives and indexicals pose a problem for Frege's theory of sense and reference, I did not mean to reject Frege's insight what when we think about an object or refer to one, some "mode of presentation" is involved. I denied only that such modes of presentation are involved in the propositions believed or expressed, which I is where Frege's theory locates them. This location, and the rest of Frege's theory, impose certain conditions on modes of presentation, which I denied that they had to meet. This is not to deny Frege's insight, but a specific theory he used to explain and accommodate it. In a sense, most of the essays in this book are attempts to find the right place, or places, to put modes of presentation, once we also grant the insights of Kaplan, Kripke, Donnellan, and others that they are not part of the propositions expressed by statements involving indexicals, demonstratives, and names" (Perry 2000, p. 21).

[40] Some philosophers treat modes of presentation as a part of the content of first-person thoughts, others treat them as a part of something else, e.g. the mental state that has the content, the way in which the content is apprehended, the mental file that is a part of the vehicle of the thought, etc. I will address this issue in more detail shortly, when I apply this generic framework to second-person thought.

[41] Canonically in Perry 1977, 1979 and Lewis 1979.

My use of the term 'mode of presentation' is intended to be taken in much the fashion that Perry describes towards the end of this quote. However, I'm not assuming that Frege's theory cannot be suitably modified such that it accommodates the insights of the "New theory of Reference." For current purposes, modes of presentation might be an aspect of the propositional content of a thought, or they might be something else related to thought, but not, strictly speaking, a part of its content.

The generic account of second-person thought unfolds in much the same fashion. The key difference is that whereas the first-person mode of presentation is already tolerably familiar to most philosophers, the second-person mode of presentation requires further clarification. Accordingly, I'll stipulate the following gloss on the *second-person mode of presentation*: it represents its object as an agent with whom I am interacting.[42] I call this a *gloss* because it clarifies what is distinctively "second-personal" about this mode of presentation, but it does so without endorsing any account of the nature of modes of presentation more generally. Further, a *second-person thought* is a thought that represents its object via a second-person mode of presentation.

I opened the chapter with an example of second-person thought (*Airport*). Here are two further examples of second-person thought:

> *Bike Path*: Imagine walking on a trail that is shared by cyclists and joggers. Without warning a cyclist flies past you. Without saying anything, you have a thought about the cyclist, which you might have expressed by yelling "you need to slow down."
>
> *Gossip*: Imagine that you and I are in the hall gossiping about someone. There are other people around us, but they seem to be going about their business, and not paying any attention to our conversation. However, I shortly realize the object of our gossip is right down the hall, hearing every word that we've said! As I look up, I see him glaring at us. I realize, as I might put it if I were talking to this person, that "you heard what we said."

While each of these thoughts might be aptly expressed by using the second-person pronoun 'you', that is not what makes them a second-person thought. Rather, it is how a thought represents its object that makes it a second-person thought. An *irreducibly second-person thought* represents its object via an irreducibly second-person mode of presentation, and an *irreducibly second-person mode of presentation* is a

[42] This is not uncontroversial. However, as I argued in the previous chapter, there is good reason to prefer this account of the second-person mode of presentation to its competitors.

second-person mode of presentation whose "cognitive significance" is not capable of being captured by any combination of first- and third-person mode(s) of presentation.

Again, I am employing a generic account of second-person thought in which I have assumed a certain gloss of the second-person mode of presentation, while otherwise employing a permissive account of the nature of modes of presentation. Before we go on to examine arguments for the irreducibility of second-person thought, it is worth pausing to further consider irreducibly second-person modes of presentation. Should there be any, what form might such modes of presentation look take?

3.1.1 The Irreducibly Second-Person Mode of Presentation.

Broadly speaking, there are two ways in which one might understand irreducibly second-person modes of presentation. If modes of presentation are to be understood as a component of the propositional content of a thought, then one might endorse the Content Reading of irreducibly second-person modes of presentation. Alternatively, if modes of presentation are to be understood as somehow related to a thought, but not as a component of its content, then one might endorse the Vehicular Reading of irreducibly second-person modes of presentation. I'll examine each reading in turn.

The *Content Reading* assumes that modes of presentation are components of the content of thoughts. As such, irreducibly second-person modes of presentation are contents whose cognitive significance cannot be captured in terms of contents consisting merely of some combination of first- and third-person modes of presentation. Proponents of the content reading might treat irreducibly second-person modes of presentation as:

> • *Fregean Senses*: Suppose that the contents of thought are propositions, which consist of Fregean senses. On such an account, irreducibly second-person modes of presentation are a distinctive kind of sense whose contribution to the proposition cannot be replaced by any combination of first- and third-person senses.
>
> • *Multi-Centered Possible Worlds*: Alternatively, suppose that the contents of thoughts are centered possible worlds. According to David Lewis,[43] the content of a *de se* thought like "My pants are on fire" cannot be captured by a proposition (e.g. the set of possible worlds in which a specific individual's pants are on fire). Instead, this content of first-person thoughts involves self-ascribing a property, e.g., PANTS ON FIRE, which is captured not merely by possible worlds, but by world-individual pairs. Thus, the content of a thought like "my pants are on

[43] See Lewis 1979, and Feit 2008 for discussion.

fire" is captured by a set of centered possible worlds, e.g. {<w,x> : x's pants are on fire in w}. Should analogous considerations push the Lewisian to use *multi-centered worlds* to capture the content of second-person thoughts,[44] it provides a means of understanding irreducibly second-person modes of presentation as a kind of unstructured thought content. Specifically, irreducibly second-person modes of presentation will involve appealing to possible worlds that are centered upon both the subject of the thought, the object of the thought, and a source of knowledge by which the subject comes to know about the object.

Alternatively, the *Vehicular Reading* denies that modes of presentation are properly understood as a component of the content of a thought. Nonetheless, modes of presentation are relevant to capturing the "cognitive significance" of thought, as is evidenced by the explanatory and epistemic work that modes of presentation are put to. This reading is often proposed by claiming that propositional attitudes, like beliefs, desires and intentions, are three place relations between a subject, a proposition, and a mode of presentation. This leaves open a variety of ways in which proponents of the vehicular reading might treat irreducibly second-person modes of presentation, including:

- *Ways of* ϕ*-ing*: Suppose that modes of presentation are ways of ϕ-ing a proposition (e.g. believing, knowing, desiring, etc.).[45] For example, suppose that Fred believes that "I am Hungry" and Sarah believes that "You are Hungry." On this account, they both believe the same thing (same proposition), but they believe it in different ways. Maybe Fred believes the proposition in a first-person way, whereas Sarah believes the proposition in a second-person way. On this account, an irreducibly second-person mode of presentation is a way of ϕ-ing which does not consist in some combination of first- and third-person ways of ϕ-ing.[46]

- *Mental States*: Alternatively, one might hold that belief states are individuated in a more fine-grained manner than the contents of these states. For example, John Perry distinguishes belief states from the contents of the propositions that we believe. He says, "we use sentences with indexicals or relativized propositions to individuate belief states, for the purposes of classifying believes in ways useful for explanation and prediction. That is, belief states individuated in this way enter into our commonsense theory about human behavior and more sophisticated theories emerging from it (Perry 1979, p. 18). On this account, the belief state that we express by saying "I am hungry" is different belief states from the one expressed by saying "the author of this paper is hungry." Nonetheless, as I am the author of this paper, the content of the two states is the same. A similar story could be told using the second person pronoun 'you'.[47]

[44] Lewis 1983 introduces the notion of a multi-centered world, and Ninan 2010, 2013 further develops it.

[45] Defenders of this position are often called Naïve Russellians, e.g. McKay 1979, Salmon 1986 and Thau 2002.

[46] An account of exactly what ways of ϕ-ing are, and how they are to be individuated, is to be provided by the proponents of this account. However, they might plausibly appeal to mental states or mental files to do such work. My presentation here is not intended to suggest that these varieties of the vehicular reading are mutually exclusive.

[47] Some subtlety is needed here, as there is more to being a second-person thought than merely being aptly expressed by 'you'. See Chapter 4, section 4.2 for some preliminary thoughts about how to tell such a story.

• *Mental Files*: Finally, one might claim that modes of presentation are mental files.[48] Mental files are the vehicles of singular thoughts, whose function is to store information about the referent of the file. Further, there are different types of files, some of which correspond to first-person thought (e.g. self files), and others of which correspond to third-person thought. On this account, an irreducibly second-person mode of presentation corresponds to a distinctive type of file, which is a different type of file from the files corresponding to first- or third-person thoughts.

For the purposes of this paper, my use of "irreducibly second-person thought" will remain neutral between the content and vehicular readings. Of course, one need not remain neutral between these proposals to accept the arguments of the next two sections. Rather, I remain neutral so as to offer a generic account of second-person thought, which aims to accommodate the different ways in which philosophers understand arguments for the irreducibility of personal perspective in thought. Suppose that one endorses the irreducibility of first-person thought by way of accepting one (or more) of the commonly-endorsed arguments of Hector-Neri Castaneda, John Perry or David Lewis. Presumably there is some way, or set of ways, in which one takes these arguments to work, and some corresponding way (or set of ways) of understanding the irreducibly first-person mode of presentation that these arguments turn on. The next two sections are meant to establish a conditional: if the irreducibly of first-person thought is shown by the kinds of arguments offered by Castaneda, Perry and Lewis, then the irreducibly of second-person thought is shown by the analogous arguments that I will offer. The generic account of second-person thought is meant to allow philosophers to accept this conditional, despite differences in their preferred account of mode of presentation.

3.2 Argument 1: Explaining Action and Interaction.

Each of the next two sections reviews a well-known argument for irreducibly first-person thought, clarifies its structure, and develops an analogous argument for irreducibly second-person thought. In doing so, I show why two commonly-endorsed reasons for endorsing the irreducibility of first-person thought can be extended to offer analogous reasons for endorsing the irreducibility of second-person thought.

Both arguments are drawn from a widely-endorsed tradition of arguments for irreducibly first-person thought which we might call the "Castaneda-Perry-Lewis tradition" (or CPL tradition). This tradition

[48] E.g. Recanati (2012). See Chapter 5 for an extended discussion of the second-person mode of presentation as a kind of mental file.

argues that the irreducibility of first-person thought can be demonstrated by acknowledging the distinctive work done by irreducibly first-person thought in explaining action, and making sense of the distinctive epistemic features of self-knowledge. For example, consider John Perry's argument for irreducibly first-person thought:

> *Sloppy Shopper*: "I once followed a trail of sugar on a supermarket floor, pushing my cart down the aisle on one side of a tall counter and back the aisle on the other, seeking the shopper with the torn sack to tell him he was making a mess. With each trip around the counter, the trail became thicker. But I seemed unable to catch up. Finally it dawned on me. I was the shopper I was trying to catch…I believed at the outset that the shopper with a torn sack was making a mess. And I was right. But I didn't believe that I was making a mess. That seems to be something I came to believe. And when I came to believe that, I stopped following the trail around the count, and rearranged the torn sack in my cart. My change in beliefs seems to explain my change in behavior." (Perry 1979, p. 3)

Let T_1 be the time at which Perry begins following the trail, and T_2 be when Perry realizes that he himself is the sloppy shopper. Shortly after T_2, Perry fixes his sack of sugar. The argument proceeds in two steps. First, it establishes a difference (in either content or mode of presentation) between Perry's belief at T_1 and his belief at T_2. Having done so, it leverages this difference to establish the irreducibility of Perry's first-person belief at T_2.

The first step of the argument might be formalized as (1)-(7):

> (1) At T_1, Perry has a belief expressible by him as "The shopper with a torn sack is making a mess."
> (2) At T_2, Perry forms a belief expressible by him as "I am making a mess."
> (3) Thus, Perry's belief at T_1 and T_2 attribute the same property to the same object.

The precise formulation of the remainder of the argument depends upon one's preferred account of mode of presentation. Suppose that (3) is not sufficient for Perry's two beliefs having the same content. Then:

> (4) If two beliefs have the same content, then they rationally explain the action actions (or dispositions to act).
> (5) Perry's belief at T_1 doesn't rationally explain his action shortly after T_2.
> (6) Perry's belief at T_2 does rationally explain his action shortly after T_2.
> Therefore,
> (7) Perry's beliefs at T_1 and T_2 don't have the same content.

The second step of the argument proceeds as (8)-(11):

(8) Perry's belief at T_1 represents its object via a third-person mode of presentation.
(9) Perry's belief at T_2 represents its object via a first-person mode of presentation.
(10) If the first-person mode of presentation of Perry's belief at T_2 is reducible to the third-person mode of presentation, then Perry's beliefs at T_1 and T_2 must have the same content.
Therefore,
(11) Perry's belief at T_2 represents its object via an irreducibly first-person mode of presentation.

Philosophers who deny that mode of presentation is a part of the content might suppose that (3) is sufficient for showing that Perry's beliefs have the same content. If so, they might substitute in the following premises into the previous argument:

(4*) If two beliefs which have the same content represent this content via the same mode of presentation, then they rationally explain the same actions (or dispositions to act).
(7*) Perry's beliefs at T_1 and T_2 don't have the same mode of presentation.
(10*) If the first-person mode of presentation of Perry's belief at T_2 is reducible to the third-person mode of presentation, then T_1 and T_2 have the same mode of presentation.

Accordingly, whether one or not one takes the irreducibly first-person mode of presentation to be a part of the content, Perry's thought at T_2 is an irreducibly first-person thought.

The most contentious premises in the argument are (4) or (4*), and (10) or (10*). But while it is contentious whether to endorse the asterisked version of the premise or not, it is much less contentious that one should endorse either (4) and (10) or (4*) and (10*), depending upon one's assessment of the implications of (3). Regardless of what one takes modes of presentation to be, if two of our *otherwise similar* thoughts employ the same mode of presentation, they ought to rationalize the same actions. Thus (4) or (4*) is true. Further, Perry's thought at T_1 and T_2 don't seem to have any relevant dissimilarities, thus either (10) or (10*) seems to be true.

The argument for irreducibly second-person thought proceeds in a similar manner. However, instead of playing an ineliminable role in the rational explanation of action, irreducibly second-person thought plays an ineliminable role in the rational explanation of interaction. To see why, it helps to distinguish clearly cases that involve mere objects, agents, and interlocutors. Thus, I'll approach the argument for irreducibly second-person thought slowly, by introducing a few contrast cases that involve a shopper named Otis.

> *Runaway Cart*: Otis is walking his cart down aisle 8 of the grocery store. It is a narrow aisle, just wide enough for two carts to pass through at a time. Over the loudspeaker Otis hears the manager yell: "If the runaway cart on aisle 8 isn't avoided, it will crash into the man on aisle 8." Knowing the manager to be trustworthy, and that the store aisles are narrow, Otis infers that unless the man on aisle 8 gets out of the way, the cart will crash into him.

As it happens, I am Otis, the man on aisle 8. When I look up and see the cart careening towards me, I reach two realizations: if *that* cart isn't avoided, it will crash into *me*. Combined these explain my reaction to the cart. Per Perry's argument, my reaction to the cart is explained by my realization that, as I would put it, "I've got to avoid that cart." But avoiding the cart requires not merely that I act, but that I act appropriately. If the cart is on my left, I must move to my right, and vice versa. The action I take is merely a reaction to a cart,[49] but it is a reaction to *that* cart, a particular cart that I am perceptually acquainted with. Hence, in *Runaway Cart*, it is my irreducibly first-person thought, alongside a thought that employs a perceptual demonstrative, which explains my action.

Next, consider a slightly different scenario. In *Buckethead*, instead of a runaway cart, there is a woman running her cart down aisle 8. This woman has a bucket covering her head, and obviously cannot see anything. As she runs down the aisle towards me, I appreciate both that she is an agent in control of the cart, and that we will crash unless I avoid her trajectory. I also appreciate that she cannot see me, and thus to avoid the crash I must react to her. Again, avoiding her cart requires not merely that I act, but that I act appropriately. If her cart is on my right, I must move to my left, and vice versa. Although my thought might arguably recognize her as an agent, and be aptly expressed by saying "…that woman" or even "…her," there is no reciprocal interaction here. Though my actions are done in response to hers, her actions aren't done in response to mine. Thus, despite the presence of another agent, the rational explanation of my action in *Buckethead* in quite similar to the explanation of my action in *Runaway Cart*.

Contrasting *Runaway Cart* and *Buckethead* with cases that involve genuine reciprocal interaction between agents uncovers the explanatory role of irreducibly second-person thought. To do so, suppose that you are the woman running her cart down aisle 8, without a bucket covering your head.

[49] In the sense of interaction specified in section 2.2, Otis can't *interact* with the cart (assuming that he recognizes it as a cart, e.g., a non-agent).

Grocery Store: Otis is walking his cart down aisle 8 of the grocery store. It is a narrow aisle, just wide enough for two carts to pass through. Over the loudspeaker Otis hears the manager yell: "If the woman running her cart down aisle 8 doesn't maneuver, she will crash into the man on aisle 8." Knowing that the manager is trustworthy, and that the store aisles are narrow, Otis infers that unless the man on aisle 8 gets out of the way, the cart will crash into him.

I look up and see the woman and her cart careening towards me. At this point, I realize that I am the man on aisle 8, and that I've got to do something to avoid a crash. But I can't act unilaterally—to avoid the crash I must avoid your cart, and you mine. You and I must coordinate our actions, but we have no time to talk. I look up and see you looking at me, trying to determine what to do. Looking at you, I think that unless *I* perform a coordinated maneuver with *you*, we will crash. I glance to the right, intending to signal my intention to turn that way. We each veer to our right, and miss each other.

Per Perry's argument, an irreducibly first-person thought explains my action in each case. But something further happens in *Grocery Store* because I appreciate that my action alone does not avoid a crash. Instead, avoiding a crash requires that you and I coordinate our actions. Due to the narrow aisles, if you veer one way, I must veer the other way to avoid you. Explaining the coordinated aspect of this interaction requires not merely that I coordinate my actions with an object (*that woman*), but with a recognized a locus of agency (*you*). Let T_1 be the time at which Otis believes that managers announcement, let T_2 be when Otis realizes that he himself is the man on aisle 8, and let T_3 be the time at which Otis appreciates the coordinated nature of the interaction. The argument is familiar:

(12) At T_2, Otis has a belief expressible by him as "unless I perform a coordinated maneuver with the woman running her cart, we will crash."
(13) At T_3, Otis forms a belief expressible by him as "unless I perform a coordinated maneuver with you, we will crash."
(14) If two of an agent's beliefs have the same content, then they rationally explain the same roles that the agent plays in an interaction (or the agent's dispositions to play the same roles in an interaction).
(15) Otis's belief at T_2 doesn't rationally explain his role in the interaction shortly after T_3.
(16) Otis's belief at T_3 does rationally explain his role in the interaction shortly after T_3.
Therefore,
(17) Otis's beliefs at T_2 and T_3 don't have the same content.
But,
(18) Otis's belief at T_2 represents its object via a first- and third-person mode of presentation.
(19) Otis's belief at T_3 represents its object via a second-person mode of presentation.
(20) If the second-person mode of presentation of Otis's belief at T_3 is reducible to the first- and third-person mode of presentation, then Otis's beliefs at T_2 and T_3 must have the same content.
Therefore,
(21) Otis's belief at T_3 represents its object via an irreducibly second-person mode of presentation.

Again, depending upon one's preferred account of modes of presentation, philosophers might substitute the following premises into the previous argument:

> (14*) If two of an agent's beliefs which have the same content represent this content via the same mode of presentation, then they rationally explain the same roles that the agent plays in an interaction (or the agent's dispositions to play the same roles in an interaction).
> (17*) Otis's beliefs at T_2 and T_3 don't have the same mode of presentation.
> (20*) If the second-person mode of presentation of Otis's belief at T_3 is reducible to the first- and third-person mode of presentation, then Otis's beliefs at T_2 and T_3 must have the same mode of presentation.

It is tempting to think that the argument for irreducibly second-person thought faces a problem that Perry's argument does not. After all, many philosophers have an easy time accepting (5) and (6) as intuitively true, but find the analogous claims made in (15) and (16) puzzling, or unclear. That is, one might initially worry that the rational explanation of the interaction in *Grocery Store* is not interestingly different from the explanation of the 'interaction' in *Buckethead*. The contrast cases are meant to clarify the similarity between the two arguments. After all, the rational explanation of Otis's reaction in *Runaway Cart* clearly seems to be different from the explanation of the interaction in *Grocery Store*. Thus, by showing that the rational explanation of the "interaction" in *Buckethead* is of the same type as the rational explanation of the reaction in *Runaway Cart*, we can see the difference in what is required to give a rational explanation of the genuine interaction that we find in *Grocery Store*, and a mere reaction (even a mere reaction to an agent).

In this section, I've belabored the structure of the two arguments to make a point. Producing two nearly identical arguments clarifies why the reasons that support accepting Perry's argument for irreducibly first-person thought also support accepting my argument for irreducibly second-person thought. Further, the contrasting cases (*Runaway Cart* and *Buckethead*) show why we cannot reduce away essentially second-person-indexical thought. Collectively, this yields a choice: either accept that "Perry-style arguments" establish the irreducibility of both first- and second-person thought, or deny that such arguments establish the irreducibility of either kind of thought.

3.3 Argument 2: Self-Knowledge and Knowledge of Others.

Nor is the phenomenon specific to the explanation of action and interaction. There are epistemic considerations that support the irreducibility of first- and second-person thought as well. David Lewis offers a second well-known argument for irreducibly first-person thought:

> *Two Gods*: Consider the case of the two gods. They inhabit a certain possible world and they know exactly which world it is. Therefore they know every proposition that is true in their world. Insofar as knowledge is a propositional attitude, they are omniscient. Still, I can imagine them to suffer ignorance: neither one knows which of the two he is. They are not exactly alike. One lives on top of the tallest mountain and throws down manna; the other lives on top of the coldest mountain and throws down thunderbolts. Neither one know whether he lives on the tallest mountain or on the coldest mountain; nor whether he throws down manna or thunderbolts (Lewis 1979, pp. 520-521).

These two gods are omniscient, in the sense that they know all the true propositions about their world. For the purposes of Lewis' argument, propositions are taken to be equally accessible to each being. Thus, "the god on the tallest mountain throws manna" expresses a proposition that is accessible to both gods, but so does "I am the god on the tallest mountain." Despite this form of omniscience, neither god knows who she is. To gain this self-locating knowledge, the gods must know more than merely all of the true propositions about the world. In addition, they must have irreducibly first-person thoughts about their location.

Putting the argument in more general terms, we might say that the gods are omniscient in that they know all of the true third-person thoughts about the world them live in, and which world it is. Third person thoughts deploy the third-person modes of presentation, e.g. thoughts that could be entertained by any omniscient god in the world, regardless of their location or properties. Despite having all of this knowledge, each god seems to exhibit a kind of ignorance of who they are. In what might this ignorance consist? Due to their omniscience, it must consist in something other than third-person thought. Because knowledge of who one is, is plausibly an instance of first-person thought, the argument shows that first-person thought cannot be reduced to third-person thought.

To extend this defense of irreducibility to second-person thought, consider a second case:

> *Dancing Gods*: Three gods inhabit an otherwise empty universe. Each god is perfectly spherical orb of light, which lacks a face and an orientation. But the gods aren't exactly alike. One god is blue, and the other two are red. While there isn't much to do, they can communicate

> telepathically, and they can "dance" with each other, which involves asking a partner to dance, and then rotating around each other in synchrony. The blue god asks a specific red god to dance, and they begin to do so. But the other red god moves in an exactly symmetrical manner on the "other side" of the blue god. Each god is omniscient in that they know which god they are, and all of the true propositions about their universe. Thus, the blue god knows, as she might put it, "I am the blue god." She also knows that "I am dancing with the god that I asked to dance." Nonetheless, I can imagine her to suffer from a kind of ignorance. She doesn't know, as she might put it to her dancing partner, which god *you* are.

In this case, there is a fact of the matter about which god the blue god is dancing with. And the blue god knows a fair amount about herself and the other two gods. She knows where she is relative to the other gods, where the other gods are relative to her, and to each other, how each god is moving, etc. She even knows, as she might put it, that "I am dancing with you." What she doesn't know is which god *you* are.

Putting the argument in general terms, we might say that the god's are omniscient in that they know all of the true third-person thoughts about the world them live in, and which world it is, alongside all of the true first-person thoughts about who they are. Despite having all of this knowledge, the blue god seems to exhibit a kind of ignorance of who she is dancing with. In what might this ignorance consist? Due to her omniscience, it must consist in something other than any combination of first- and third-person thought. Because knowledge of who you are, is plausibly an instance of second-person thought, the argument shows that second-person thought cannot be reduced to any combination of first- and third-person thought.

One might worry that spatial information cannot be so easily excluded. The example assumes that the relevant kind of spatial information is relational. But if the space-time of the world has an objective structure (e.g. a coordinate system with an objective origin), it would cause problems for the example. The blue god might then know things like "I am at 3, 4" and "I asked the red god at 4,5 to dance." But surely it is possible for space-time to have this relational character, and this is all that the example needs.

More plausibly, one might worry that the blue god could simply suppose that such a coordinate system were in place, and then, due to her omniscience, know all of the relevant counterfactuals of the form "If coordinate system C were true, and if I were at X,Y, then the god that I asked to dance would be at X_1,Y_1." The coordinate system being supposed would give the universe a default orientation (despite

the fact that the gods lack one), and would give the blue god a way to differentiate between the two red gods. However, this doesn't alleviate the blue god's ignorance of which god she is *actually* dancing with. Because the universe doesn't have the objective coordinate system being supposed, the blue god cannot know which god she is dancing with by knowing which god should would be dancing with should different conditions apply.

A second objection gets closer to the heart of the matter. This objection contends that the blue god might name the two red gods (R_1 and R_2), and thereby infer which god *you* are. The basic idea is that the blue god has a quick and easy answer to the question which god am I dancing with? It is: "I am dancing with R_1" and "I am not dancing with R_2." Since you (as the god might think it) are the one with whom I am dancing, the blue god can thereby infer who you are.

This objection fails, as the inference doesn't go through. While the blue god knows things like "I am dancing with R1," there is still a sense in which she doesn't know who she is dancing with. To see why, it helps to consider how the analogous response fairs as a reply to *Two Gods*. Suppose that the two gods are named G_1 and G_2. G_1 therefore knows things like "G_1 is the god in the valley" and "G_2 is the god on the mountain." Granting this, there nonetheless seems to still be a sense in which these gods remain ignorant. Why? Because although each god can think about herself in third person terms, it isn't until she can think about herself *as herself* that she dispels the relevant kind of ignorance. Thus, it isn't enough to be able to think about the right object. Instead, one must think about the right object *in the right way* (or under the right mode of presentation). The gods can't infer this self-conscious thought from the (merely third-personal) things that they already know.

The same basic reply works for *Three Dancing Gods* as well. The blue god can think about each of the other gods, and even think about the god with whom she is dancing. But, in order to dispel the kind of ignorance in question, the blue god has to think about her dance partner in a second-person manner. Therefore, all appearances to the contrary, the blue god can't infer (as she might express it) "You are the god I am dancing with" from what she already knows. Knowing that "I am dancing with R_1" and "I am dancing with you" ought to seem importantly different to philosophers. Why? Because I might not know that "you are R_1." Of course, the blue god can think about R_1 in first- and third-person terms, but it isn't until she can think about R_1 *as an agent with whom I am interacting* that she dispels the relevant kind of ignorance.

Finally, one might worry that this example is simply too strange to learn much from. After all, *Dancing Gods* effectively takes away all of the information about spatial relations that one might otherwise use to think about the objects in the external world (e.g., relations the other objects in the universe, and the intrinsic orientation that a body affords). In this respect, it is quite different from *Two Gods*. Given that we are embodied beings, whose perceptual and communicative apparatus seems to be spatial through and through, it might seem wrongheaded to think that we could draw a verdict about second-person thought from such a strange example. However, one might instead understand the example as eliminating the temptation to think about interaction in merely spatial terms. After all, if *Dancing Gods* is possible, then the second-person mode of presentation is not identical to any combination of first- and third-person modes of presentation. This remains true even if we can't pull apart the conditions in which we apply these types of modes of presentation in the actual world. Instead, *Dancing Gods* takes the information about spatial relations away, and contends that such gods are still able to think about each other both as agents, and as partners in interaction. Thus, the case isn't obfuscating the role played by spatial representation in cognitive systems like ours, but highlighting the further role played by second-person thought.

All said, one might remain unconvinced by *Dancing Gods*. But I'm not claiming that one must be. My claim is that the arguments yielded by *Two Gods* and *Dancing Gods* have analogous structures. One is free to reject both arguments. What I've argued is, in effect, that if one is convinced of the irreducibility of first-person thought by *Two Gods*, one ought also to be convinced of the irreducibility of second-person thought by *Dancing Gods*.

3.4 Local Objections to Irreducibly Second-Person Thought.

The previous two sections illustrate how one might extend commonly endorsed arguments for the irreducibility of first-person thought into arguments for the irreducibility of second-person thought. I've crafted the analogous argumentative considerations operative in each case, thus illustrating why the arguments stand and fall together. This conclusion might be taken in two ways. For philosophers who already accept the irreducibility of first-person thought on the grounds provided by the CPL-tradition, I have effectively argued for the irreducibility of second-person thought. But, for those philosophers who reject the arguments of the CPL-tradition, my arguments for irreducibly second-person thought will inherit their flaws. My thesis is, therefore, a conditional claim:

> **(C)** if the irreducibly of first-person thought is established by the kinds of arguments offered by the CPL tradition, then the irreducibly of second-person thought is established by analogous arguments.

As usual, there are two ways to reject the consequent of a conditional. *Local Objections* accept that CPL arguments establish the irreducibility of first-person thought, but nonetheless reject my arguments for irreducibly second-person-indexical thought. Such objections accept the antecedent of the conditional, and deny that the consequent follows. *Global Objections* reject both the CPL arguments for irreducibly first-person thought, and my proposed extensions. These objections deny the antecedent of the conditional (C), and thus deny the truth of its consequent, despite the truth of the entire conditional. I'll consider each form of objection in turn.

The dominant source of philosophical skepticism about irreducibly second-person thought stems from local objections. Many philosophers accept irreducibly first-person thought, often on the argumentative grounds put forth by the CPL-tradition. Nonetheless, irreducibly second-person thought enjoys no such popularity. There are two basic options for defending this asymmetrical between first- and second-person thought: establish some relevant difference between the first- and second-person mode of presentation, or show that while irreducibly second-person thought can be reduced, irreducibly first-person thought cannot.

The main attempt to establish a relevant difference asserts that second-person phenomena are fundamentally linguistic in nature. Thus, irreducibly second-person thought, but not irreducibly first-person thought, confuses features of the language we use to express a thought for features of the thought expressed by that language. Richard Heck makes the point vividly:

> "I mean to deny that there is any such thing as an essentially indexical second-person belief. The phenomenon of the second-person is a linguistic one, bound up with the fact that utterances, as we make them, are typically directed to people, not just to the cosmos…The word 'you' has no correlate at the level of thought." (Heck 2002, p. 12)

Joining Heck in this assessment, Kent Bach adds that for the second-person pronoun 'you', "although there is a linguistic rule that determines its reference, relative to a context, as the addressee, there is no corresponding second-person way of thinking of someone." (Bach 2007, p. 416) Put together, this yields an argument: the linguistic meaning of 'you' is fixed by a rule that makes reference to address (e.g., 'you' means "the person addressed by the speaker of this utterance"), and there is no analog of

address at the level of thought. Thus, there is no distinctive mode of presentation, expressible by 'you', to be deployed in second-person thought.[50]

This is an important argument, but it simply doesn't apply to the account of irreducibly second-person thought I've defended here. As I understand it, the second-person mode of presentation is understood in terms of its function, and not in terms of the meaning of the second-person language that may (or may not!) be used to express it. The function of second-person thought involves recognizing another agent as an agent with whom I am interacting. While there is surely more to be said by way of spelling out this function, there is no reason to suspect that such an account will run afoul of this language-based objection. Second-person thought is no more or less a "linguistic" phenomenon than first-person thought.

More pressing local objections stem from the attempt to reduce second-person thought to first- and third-person thought. The most prominent attempts to do this assume that there is irreducibly first-person thought, and then attempt to reduce second-person thought some combination of first-person thought alongside either descriptivism or perceptual demonstratives. To show why neither of these approaches work, I'll examine two accounts of second-person thought: one that stems from a simple form of descriptivism and another from Christopher Peacocke.[51]

It helps to begin with a simple version of the descriptivist objection. John Perry suggests this objection when he noted that "it is plausible to suppose that other indexicals can be eliminated in favor or 'I' and 'now'." (Perry 1979, p. 16) Superficially, Perry might be read as suggesting that we can replace the second-person mode of presentation with some combination of non-indexical terms, and the first-person modes of presentation affiliated with indexicals like 'I' and 'now'. The proposal is problematic, as the most natural way to convert a second-person mode of presentation into a conjunction of other modes of presentation is to create something akin to a definite description. But the content of some indexicals, including 'you', appears to be different from the content of a definite description, like 'the

[50] Salje 2017 dubs this the "argument from address" and discusses it at greater length. See Haddock 2014 and Salje 2017 discussions of the nature of address and its bearing upon this argument.

[51] Both Lewis and Peacocke endorse content readings of second-person thought. Thus they aren't attempting to show that no version of my arguments for irreducibly second-person thought might succeed. But I contend that the arguments of section 2 and 3 hold generally, e.g., for both the content and vehicle versions of the thesis. Thus, it is important (and informative) to show why Lewis and Peacocke's attempts at reduction don't succeed.

person that I am now talking to'. To see why, imagine expressing an second-person thought about one of the many people you are simultaneously talking to. If the thought is aptly expressed using 'you', it remains unclear how to express this thought by using 'I' and 'now' alongside other non-second-person words. Absent more satisfactory ways of translating the apt expression of second-person thought into 'I', 'now', etc., the proposed reduction will fail. Accordingly, the proposed descriptivist reduction of second-person thought will fail for analogous reasons.

Christopher Peacocke contends that second-person thoughts can be reduced, and offers a bit more of an account. Peacocke grants that "there is a use of 'you' in thought, and that it can occur in thinking that is not expressed out loud in language, and is not intended as part of any communicative act" (Peacocke 2014, p. 247). Thus he isn't merely objecting to second-person thought on the linguistic grounds rejected earlier. But he contends that the concepts, or identifying conditions, that make up the content of a second-person thought are nothing more than identifying conditions that correspond to 'I' and 'here' alongside a perceptual demonstrative in the apt expression of a thought. He rejects the idea that there is "distinctive second-person concept" and instead claims that "any occurrence of 'you in thought can equally be explained in terms of other concepts that do not involve an irreducible second-person concept." (*ibid*, p. 247) Peacocke has in mind the following picture: the occurrence of the second-person pronoun 'you' in the apt expression of a thought can be treated as a perceptual demonstrative (e.g. expressing *that shopper*, or *that cyclist*) in a complex expression that involves first-person concepts. Of course, the treatment of these perceptual demonstratives might in turn involve essentially indexical content, but any essentially indexical components of thought will correspond to irreducibly first-person modes of presentation, like those expressed by 'I', 'here' and 'now.' Thus the second-person mode of presentation might be reduced to its first and third-person elements.

However, Peacocke's proposal fails to reduce the second-person mode of presentation, as it cannot distinguish Otis's thoughts about the runaway cart, the agent with the bucket over her head, and the agent with whom he is engaged in interaction. Otis merely reacts to the runaway cart and the buckethead, but interacts with the other agent in *Grocery Store*. His action is done in response to hers, hers in response to his, etc. Explaining this reciprocal interaction requires appeal to irreducibly second-person thoughts because an interaction involves each agent recognizing the other as *an agent with whom I am interacting*. It is this recognition of another agent that captures the difference between merely reacting to an agent (e.g., the buckethead), and full-blown interaction. And it is this form of recognition

that, I've argued, is characteristic of the irreducibly second-person mode of presentation. Thus, a mental representation that is richer than a mental demonstrative, or even a complex expression consisting of a demonstrative alongside first-person concepts, is needed to capture the second-person mode of presentation. Attempts to reduce second-person thought by appeal to descriptions and demonstratives seem to fail, and fail for familiar reasons.

The arguments for irreducibly second-person thought that I've given are natural extensions of arguments that are already widely accepted. I've further argued that there aren't reasons to think that second-person thought in particular fails or reduces away to first- and third-person thought. To this, I'll add a conjecture: should any attempt to reduce second-person thought succeed, analogous maneuvers can also be utilized to reduce first-person thought.[52] If true, this poses a dilemma for any seemingly local objection: either it isn't genuinely local, or it fails to adequately reduce the second-person perspective. Until a local objection can split this dilemma, (C) holds.

3.5 Towards a Basic Argument for Irreducibly Second-Person Thought.

Of course, one needn't deny (C) to reject irreducibly second-person thought. *Global Objections* deny that arguments from the CPL-tradition succeed in establishing the irreducibility of first-person thought. Proponents of such objections might accept (C) as vacuously true due to its false antecedent, and thus conclude that I've given no reason to accept the irreducibility of second-person thought.

Most global objections hinge on wide-ranging disagreements about the nature of content, the relation between thought and language, and the explanatory role of mental content. This makes it difficult to offer a straightforward response to global objections, as it would require assessing the proposed treatments of a variety of related issues. For example, Herman Cappelen and Josh Dever contend that arguments for "essentially indexical thought" bunch together fragments of other phenomena, and make the mistake of "thinking that there's something—call it essential indexicality—that ties all of these fragments together" (Cappelen and Dever 2013, p. 3). As such, Cappelen and Dever reject the arguments of the CPL-tradition, and deny that the first-person perspective plays any important role in rationalizing action, or in explaining self-knowledge. Ruth Millikan offers a related objection,

[52] Some varieties of "token-reflexive" accounts of indexical content might fall into this category. Such attempts might succeed in reducing second-person thoughts to first-person thoughts plus demonstratives, but, if successful, such accounts could also treat first-person thoughts, like I am F, as having the following content: the thinker of this thought is F.

arguing that it is not the semantics of "essential indexical thoughts" that distinguishes them, but their psychological role. More specifically, she argues that "there can be no thought that has the (Kaplan-style) character of 'I'" (Millikan 1990, p. 724). Further, Michael Devitt and Ofra Magidor have each rejected the claim that *de se* attitudes pose any particularly thorny philosophical problem.[53] Each of these authors denies that the arguments in the CPL tradition establish the claims related to irreducibly first-person thought that they are often cited as establishing.

A response is surely owed to these global objections. But instead taking up this argumentative burden now, for current purposes I'll explain why all parties to the debate, even global objectors, have reason to endorse (C). Appreciating why (C) is true clarifies the structure of well-known arguments for irreducibly first-person thought. Having done so, these arguments can be more accurately assessed by skeptics and proponents alike. Further, skeptics about irreducibly first-person thought stand to gain from the truth of (C). Many philosophers find irreducibly second-person thought to be less intuitively motivated than irreducibly first-person thought. If global objections are to prove compelling, the truth of (C) ought to be grist for the skeptical mill. Much as (C) might seem to many philosophers to figure as a premise in an argument for irreducibly second-person thought, in the skeptics hands, it might be used to argue against the irreducibility of first-person thought.

Operating in the background of this paper is a kind of basic argument for irreducibly second-person thought. The irreducibility of first-person thought has been shown by a set of widely celebrated arguments stemming from the CPL tradition. While there is disagreement about exactly how these arguments work, and what they establish, I've offered a generally acceptable way of framing their lesson: they illustrate the irreducibility of the first-person mode of presentation. Further, I've argued that if the irreducibly of first-person thought is shown by the kinds of arguments offered by the CPL tradition, then the irreducibly of second-person thought is shown by the analogous arguments that I've developed here. What should we make of this conditional? I haven't yet said. But the generic account of second-person thought that I've proposed disentangles debates about the irreducibility second-person thought from objections driven by ancillary commitments, e.g. that the content of irreducibly first-person thought is itself "indexical" in some sense, or the thought what makes a thought an "essentially indexical thought" is that it must be expressed by some certain indexical

[53] Devitt 2013 and Magidor 2015 respectively.

language. Thus it allows debates about the irreducibility of second-person thought to proceed unencumbered by the baggage that they sometimes bring with them.

I've argued that irreducibly first- and second-person thought are of a piece: if you accept the one, you should accept the other. Like many philosophers, I endorse irreducibly first-person thought because I accept some version of the arguments for irreducibility stemming from Castaneda, Perry and Lewis. Thus, by my lights, what I've offered here is an argument for irreducibly second-person thought. The existence of such thought is applicable to an array of philosophical projects. But even setting aside these applications, irreducibly second-person thought is interesting because it clarifies the philosophically interesting features of so called "essentially indexical thought." While philosophers have rightly recognized irreducibly first-person thought as a distinctive form of thought, they have incorrectly restricted this irreducibility to the first-person (or "subjective") perspective. Instead, analogous considerations show that both first- and second-person thought are irreducible, thus shifting our focus to these two varieties of personal perspective. The result is that much as self-conscious thought might differ from thought about other objects, thought about other agents might different when interacting with them, as opposed to merely observing them.

Chapter 4: The Basic Argument for Irreducibly Second-Person Thought

Thus far, the two central conclusions that I have reached involve clarifying the second-person mode of presentation, and beginning an argument for the irreducibility of second-person thought. After stipulating some terminology and focusing upon second-person thought, *chapter 2* provides a gloss on the second-person mode of presentation. On my account, second-person thought employs a second-person mode of presentation, and this mode of presentation should be understood in terms of its recognition function: representing its object as a rational agent with whom I am engaged in interaction.

Chapter 3 began an argument for the irreducibly of second-person thought. I show why the same kinds of considerations that support the irreducibly of first-person thought can be extended to support the irreducibility of second-person thought. Doing so supports a conditional: if the irreducibly of first-person thought is shown by the kinds of arguments offered by the CPL tradition, then the irreducibly of second-person thought is shown by analagous arguments.

This conditional should be understood against a philosophical orthodoxy which acknowledges that irreducibly first-person thought exists, for (more or less) an established set of reasons. That is, many philosophers accept the irreducibly of first-person thought on the basis of some compelling examples and arguments provided by Hector-Neri Castaneda, John Perry and David Lewis and the tradition following them (which I'll collectively call the CPL tradition). Even the opponents of this tradition acknowledge that not only does philosophical orthodoxy endorse irreducibly first-person thought, but that major contemporary research programs are devoted to pursuing its implications of such thought.

When combined, this yields a *Basic Argument for Irreducibly Second-Person Thought*:

> (1) The irreducibly of first-person thought is shown by arguments in the CPL tradition.
> (2) If the irreducibly of first-person thought is shown by arguments in the CPL tradition, then the irreducibly of second-person thought is shown by analogous arguments.
> Therefore,
> (3) The irreducibly of second-person thought is shown by analogous arguments.

The *Basic Argument* is intended to be what Nancy Cartwright has called a short, stocky, and tangled argument (Cartwright 2018). The argument is short in that it consists of only few premises, and stocky in that its conclusion "covers a lot of territory" (Cartwright 2018, p. 4). The expanse of "territory covered" can be illustrated by noting that the argument doesn't hinge upon the specific claims of a theory of content or thought. As such, the *Basic Argument* can be endorsed on nearly any plausible theory of thought content, including by those who hold that the first- or second-person modes of presentation aren't properly speaking an aspect of the content of a thought at all.

More importantly, I've argued that the reasons for endorsing the premises align, in the sense that the premises seem to stand and fall together. When arguments have the same structure, unless we can find a relevant difference, we ought to evaluate the arguments in the same manner. The *Basic Argument* is tangled in that it is supported by "a rich network of interrelated arguments, each firmly attached to the ground, some with shared premises but where a great many also have a number of independent premises and…premises that come from a variety of different places outside the immediate domain in which the conclusion lives" (Cartwright 2018, p. 4). In this case, the "ground" is a set of cases, the myriad interpretations of these cases that have been used to argue for irreducibly first-person thought, and the arguments based upon these interpreted cases. The arguments of the CPL tradition identify a specific means of supporting (1),[54] namely that irreducibly first-person thought does distinctive explanatory work, and/or is needed to capture certain epistemic features of thought. By extending the same sorts of considerations to irreducibly second-person thought, the argument for the irreducibility of second-person thought proceeds on analogous grounds to those that philosophical orthodoxy uses to endorse the irreducibly of first-person thought. Thus, (2) is true as well. Because the very same sorts of considerations support both (1) and (2), one cannot obviously reject (2) without rejecting (1) as well. The *Basic Argument* is tangled indeed.

How does the *Basic Argument* fare? I've already defended (2), but an assessment of the argument requires further discussion of (1). This chapter offers a defense of (1), but it won't proceed by attempting to compel the reader to accept (1). I say this for two main reasons. First, any attempt to do so will consume the discussion of second-person thought. A full-scale defense of the irreducibly

[54] Throughout, I'll use "arguments from the CPL tradition" (and sometimes specifically "Perry-Lewis Style of Arguments") as shorthand for the familiar arguments that seek to establish the existence of irreducibly first-person thought on the basis of its distinctive epistemic and explanatory features.

of first-person thought is a project unto itself. Second, offering such an argument is inimical to my purposes here. Any argument for (1) that is detailed enough to compel even limited acceptance is likely to limit the stocky, tangled appeal of the *Basic Argument* by taking a stand on various issues related to the nature of content, the relation between language and thought, etc.

As it stands, the basic argument might be acceptable to philosophers who otherwise disagree about many related issues. This is as it should be. (1) seems to me, as it does to many philosophers, to be an eminently plausible claim, and its plausibility can be captured in a number of different theoretical frameworks. Thus, it is desirable for arguments for irreducibly first- or second-person thought to proceed at a general level, independently of any particular account of each, and I have crafted the argument accordingly. The real interest of the *Basic Argument* lies in its ability to draw the motivation for irreducibly second-person thought from one's motivations for endorsing the irreducibly of first-person thought, whatever they might be.

All that said, (1) isn't so unobjectionable that it needs no defense. In fact, it has recently faced objection in the form of critical discussion of Perry and Lewis's arguments for irreducibly first person thought—the very arguments upon which my case for (2) hinges—thus merely presupposing (1) is objectionable in the current context of discussion. This suggests a strategy: defend (1) by noting its plausibility and rejecting the arguments against it. This chapter argues for the following claim: we have reason to accept premise (1) of the *Basic Argument* because it *is* plausible, and the arguments against it *do* fail. Thus, the *Basic Argument* succeeds.

4.1 Setting the Argumentative Standards: A *Defense* of Irreducibly First-Person Thought.

Again, my strategy is this chapter is to defend (1). This is not to attempt to compel belief in (1), which would evoke standards that couldn't be met in a satisfactory manner. Nor it is merely to presuppose (1), which leaves the claim without any positive epistemic status, and apt to be dismissed. Instead, I show both that arguments from Perry and Lewis establish the plausibility of (1), granting it some default positive epistemic status [which I'll call a *presumption* in favor of (1)], and that we lack reason to reject this status. Thus, after scrutiny, (1) retains a positive status.

Why might one endorse the irreducibility of first-person thought? The CPL tradition identifies two reasons why first-person thoughts don't reduce to third-person thoughts. First-person thoughts seem

to have distinctive epistemic features that cannot be captured by other kinds of thoughts. Consider John Perry's story about Rudolf Lingens:

> An amnesiac, Rudolf Lingens, is lost in the Stanford Library. He reads a number of things in the library, including a biography of himself, and a detailed account of the library in which he is lost…He still won't know who he is, and where he is, no matter how much knowledge he piles up, until that moment when he is ready to say, "This place is aisle five, floor six, of Main Library, Stanford. I am Rudolf Lingens" (Perry 1977, p. 492).

Here is seems that despite his extensive knowledge of both the library and of the life of someone named Rudolf Lingens, Lingens lacks a kind of self-knowledge that we all (presumably) have. Lingens won't understand who or where he himself is until he has the relevant irreducibly first-person thoughts. And gaining that knowledge requires having thoughts that represent Lingens in a distinctively first-person manner.

One might worry that the need for irreducibly first-person thought is an artifact of ignorance: if Lingens simply knew more about Lingens, the need for first-person thoughts will evaporate. But while Lingens isn't omniscient, it doesn't seem that irreducibly first-person thoughts are required merely to make up for some crucial lack of factual knowledge on Lingen's part. David Lewis illustrates this with *Two Gods*:

> Consider the case of the two gods. They inhabit a certain possible world and they know exactly which world it is. Therefore they know every proposition that is true in their world. Insofar as knowledge is a propositional attitude, they are omniscient. Still, I can imagine them to suffer ignorance: neither one knows which of the two he is. They are not exactly alike. One lives on top of the tallest mountain and throws down manna; the other lives on top of the coldest mountain and throws down thunderbolts. Neither one know whether he lives on the tallest mountain or on the coldest mountain; nor whether he throws down manna or thunderbolts (Lewis 1979, pp. 520-521).

Here, the two gods are omniscient, in the sense that they know all the true propositions about their world, and yet they seem to lack much the same knowledge that Lingens did. One might object to various details about the case, e.g. Lewis's account of propositions, the notion of omniscience, etc. But it isn't clear how these objections could rid us of the lingering suspicion that there is something about themselves that Lingens and the two gods don't know,[55] and that however we characterize it

[55] Stalnaker 2016 is instructive in this regard. After famously arguing that Lewis' notion of omniscience fails, and that we can't understand the *Two Gods* case in such a way that supports irreducibly first-person thought, Stalnaker constructs his own argument for irreducibly first-person thought that sidesteps such issues.

exactly, our knowledge of who and where we are requires having the relevant irreducibly first-person thoughts.

Nor is this "self-locating knowledge" inconsequential or idle. A second line of argument suggests that irreducibly first-person thought does explanatory work which cannot be done by other kinds of thought. Most famously, our capacity for rational action requires having irreducibly first-person thoughts.[56] John Perry's *Sloppy Shopper* example (Perry 1979), discussed in the previous chapter, shows that it isn't until the shopper has the relevant irreducibly first-person thought that we can give a rationalizing explanation of the shopper's action.

This explanatory indispensability suggests some kind of connection between agency and irreducibly first-person thought. But even if one balks at the idea that rationalizing explanations of action require appealing to irreducibly first-person thought, there seem to be other indispensable explanatory roles that such thoughts play. For example, it seems like explaining what triggers a self-conscious emotion like shame requires appealing to an irreducibly first-person thought. Imagine the case of John, a reveler in Las Vegas who lost his phone during a wild night out. While looking for his phone, John returns to all of the places that he remembers visiting the night before. At each one, he hear about the antics of a man wearing the baby blue tuxedo. Having heard these stories, John thinks that the man who was wearing the baby blue tuxedo ought to be ashamed of himself. Gradually, John realizes that, as he might put it, I am the man who was wearing the baby blue tuxedo. It is only after reaching this realization that John feels ashamed of himself.

Nor is this indispensibility limited to folk psychological explanation. The ability to form and recall episodic memories, and to efficiently construct plans, each seem to require the ability to form irreducibly first-person thoughts (see Horgan and Nichols 2016 for further discussion). The recall and deployment of episodic memories seems to require not only that I remember an episode, but that I remember that it happened to me myself. Similarly, an efficient planning architecture requires tagging intentions in some way, such that I don't end up constructing conflicting intentions about future action. In these last two cases, the argument for irreducibly first-person thought still hinges upon the distinctive explanatory work that such thoughts can perform. But instead of a folk psychological

[56] Here I use thought loosely, such that it might include intentions as well as beliefs, desires, etc. See Babb 2017 for discussion of the claim that arguments in the CPL tradition ought to focus on intentions.

explanation (rationalizing action), the explanatory context is more scientific—that of designing an efficient cognitive architecture.

In some version or another, I find each of these arguments to be compelling. Of course, there are disagreements about exactly which versions of the arguments one should endorse, exactly what they show, and how they show it. But many philosophers agree that these arguments succeed in showing that irreducibly first-person thought exists when suitably interpreted. To mark this philosophical consensus, I'll say that there is a presumption in favor of irreducibly first-person thought.

But presumptions can be overridden if needed. Not everyone accepts the irreducibly of first-person thought, or that CPL-style arguments establish it. Further, these sceptics have given arguments against both of these claims. Why, then, might one override the presumption in favor of irreducibly first-person thought? I'll examine two main types of objection. First, there are alleged problems with the arguments for irreducibly first-person thought. These objections put aside questions about the existence of such thought, and instead contend that the arguments given don't (and can't) establish such claims. Second, there are objections to irreducibly first-person thought itself. These arguments object to irreducibly first-person thought directly, by arguing that such thoughts are incoherent, or otherwise require paying theoretical costs that we can't afford.

The next two sections develop instances of each type of objection. In each case, after clarifying the objection, I argue that it poses no genuine problem for the CPL style of argument, or for irreducibly first-person thought. If these replies are successful, and the presupposition in favor of irreducibly first-person thought remains undefeated, then the *Basic Argument* succeeds.

4.2 Is there a problem with the arguments for Irreducibly First-Person Thought?

A first form of objection contends that there is a problem with the existing arguments given for irreducibly first-person thought, and in particular with the arguments given by John Perry and David Lewis. This form of objection states that because Perry and Lewis's arguments in favor of irreducibly first-person thought fail, the many philosophers who cite it approvingly have no reason to accept it.

Why might these arguments fail? One objection contends that the arguments for irreducibly first-person thought are *merely* Frege cases, and thus fails to establish anything about the nature of thought itself. Instead, they establish only that action explanation contexts are opaque.

4.2.1 Perry's Argument for Irreducibly First-Person Thought.

John Perry and David Lewis have produced an intuitively-gripping set of examples involving sloppy shoppers, omniscient gods, and shabby pedagogues, which have convinced many philosophers of the irreducibility of first-person thought—which is variously called essentially indexical thought, self-locating thought, or *de se* thought.[57] While arguments from this "Perry-Lewis tradition" are widely accepted, they face staunch scrutiny. Ofra Magidor contends that irreducibly first-person thought doesn't solve any distinctive philosophical problem, and thus rejects the "myth of the *de se*" as unmotivated. More pointedly, Herman Cappelen and Joshua Dever contend that the arguments for irreducibly first-person thought systematically fail, as what have typically been "paraded as instances of essential indexicality" merely establish opacity, the failure of co-referential terms to substitute *salva veritate* in certain contexts (Cappelen and Dever 2013, p. 28).

While I reject their skeptical conclusions, this critical attention shows that the arguments in the Perry-Lewis tradition stand in need of clarification. And it is by clarifying the arguments for irreducibly first-person thought that we can show why the skeptical challenges fail. I'll limit my discussion to John Perry's example of the *Sloppy Shopper*, and Cappelen and Dever's criticism that Perry offers a *mere* Frege case, establishing only the opacity of action explanation contexts. In reply, I argue that the most plausible form of Perry's argument is a broadly empirical argument—a form of high-level "speculative psychology" of the kind Jerry Fodor advocated in the introduction to *The Language of Thought*—that proceeds in two steps. After unpacking Cappelen and Dever's criticism of Perry's argument, and situating the ensuing discussion, I clarify the two-step structure of Perry's argument for irreducibly first-person thought. While the first step of the argument might be represented as a Frege case, the second step deploys a principle that captures a contingent, but explanatory, connection between first-person thought and action. On this reading, Perry's famous example occurs in the first step of an argument, the second step of which is no mere Frege case.

[57] While tend to use the general term "irreducibly first-person thought," discussions of Perry tend to use the terminology he famously introduced "essentially indexical thought." For the sake of continuity with my sources, I often use Perry's more specific term in favor of my more general one in this section. Nothing is meant to hinge upon this choice.

4.2.2 The Objection: Mere Frege Cases.

To begin, recall John Perry's example of the sloppy shopper:

> ***Sloppy Shopper*.** I once followed a trail of sugar on a supermarket floor, pushing my cart down the aisle on one side of a tall counter and back the aisle on the other, seeking the shopper with the torn sack to tell him he was making a mess. With each trip around the counter, the trail became thicker. But I seemed unable to catch up. Finally it dawned on me. I was the shopper I was trying to catch…I believed at the outset that the shopper with a torn sack was making a mess. And I was right. But I didn't believe that I was making a mess. That seems to be something I came to believe. And when I came to believe that, I stopped following the trail around the count, and rearranged the torn sack in my cart. My change in beliefs seems to explain my change in behavior. (Perry 1979, p. 3).

At a minimum, Perry shows that action explanation contexts are opaque. Suppose that John Perry believes that the shopper with the torn sack is making a mess. He doesn't thereby believe that he himself is making a mess, even if he is the shopper with the torn sack. For he might believe that the shopper with the torn sack is making a mess the whole time, and it isn't until he believes that he himself is making a mess that his belief rationalizes fixing his sack of sugar. This further suggests a change between two different beliefs. On the basis of the opacity of action explanation, and some claims about which beliefs explain which actions, Perry concludes that certain first-person thoughts are irreducible—they cannot be reduced even to co-referential third-person thoughts—as only the first-person thought rationalizes his action.[58]

Lurking behind all the excitement about agency, Cappelen and Dever see little more than a Frege case. Frege cases establish that a given context is opaque—that co-referential terms cannot be substituted *salva veritate*. Perry's argument might be taken to establish the following thesis:

> **Indexical Opacity (IO)**: There's a set of indexicals, I-SET, that cannot be substituted *salva veritate* in action-explanation contexts by any other expressions. (*ibid.*, p. 33)

But instances of (IO) might merely be instances of (GO):

> **General Opacity (GO):** Co-referring expressions cannot be substituted *salva veritate* in action-explanation contexts. (*ibid.*, p. 33)

[58] Exactly what the failure of reduction consists in will depend upon how one thinks about the distinctive aspects of first-person thought. If one thinks that irreducibly first-person thoughts have a distinctive kind of content, then it is the content of such thoughts that is irreducible. If one thinks that irreducibly first-person thoughts represent their content in a distinctive way, then it is the way of representing their content that is irreducible. The discussion here remains neutral on this issue.

And (GO) is a generic lesson about language, it applies to indexical and non-indexical language alike. Consequentially, establishing (GO) can't support claims about the irreducibility of first-person thought. Thus, the success of Perry's argument hinges on whether it establishes (IO) for reasons specific to first-person thought, or whether it establishes (IO) merely as a specific instance of (GO). I'll call these two readings of *Sloppy Shopper* the (IO)-reading and the (GO)-reading respectively.

Cappelen and Dever offer two considerations in favor of the (GO)-reading. First, by using *Sloppy Shopper* as a template, they construct a second example which mirrors the opacity it demonstrates, but without employing indexicality.

> ***Superman/Clark Kent***: Pushing my cart down the aisle I was looking for CK to tell him he was making a mess. I kept passing by Superman, but couldn't find CK. Finally, I realized, Superman was CK. I believed at the outset that CK was making a mess. And I was right. But I didn't believe that Superman was making a mess. That seems to be something that I came to believe. And when I came to believe that, I stopped looking around and I told Superman to clear up after himself. My change in beliefs seems to explain my change in behavior. (Cappelen and Dever, 2013, p. 33)

Because the opacity evinced by *Sloppy Shopper* generalizes to contexts that don't employ indexicals, Cappelen and Dever contend that "there's nothing deeply central about indexicals here" (*ibid*, p. 33). Instead, both cases share a common abstract structure, and establish instances of (GO).

Proponents of Perry's argument might find significant differences between the two cases, especially having to do with the connections between indexicality and agency. But they are wrong, Cappelen and Dever argue, because there is no philosophically significant connection between indexicality and agency. Thus, Perry's argument stalls having merely established opacity, as there are no suitable inference principles from which to derive claims about irreducibly first-person thought from claims about the explanation of action.

To make this second point, Cappelen and Dever consider the form that such principles might take, and eliminate all philosophically interesting contenders. Perry uses examples like *Sloppy Shopper* to argue that action rationalizations are incomplete until they include a belief that is expressed using first-person indexical like 'I' or 'me*'. The rationalization of Perry's action in *Sloppy Shopper* then proceeds as follows:

> **Belief**: JP is making a mess.
> **Belief**: I am JP.
> **Belief (inferred):** I am making a mess.
> **Desire**: I not be making a mess.
> **Belief**: If I fix my sack of sugar, I will not be making a mess.
> **Action**: I fix my sack of sugar. (Adapted from Cappelen and Dever 2013, p. 36)

If the "essential indexicals" were removed, Perry argues, the action would no longer be rationalized. Thus, in order to rationalize an action we must appeal to irreducibly first-person-indexical thoughts. So understood, Perry's argument turns on the following principle:

> **Impersonal Incompleteness Claim (IIC):** Impersonal action rationalizations (IAR's) are necessarily incomplete because of a missing (first-person) indexical component. (*ibid.*, p. 36)

Though (IIC) has proponents,[59] the strength of the claim is striking. Cappelen and Dever consider theses which weaken the 'necessarily' in (IIC) to 'deeply physically necessarily', 'humanly necessarily', etc. But they ultimately reject all of these claims for the same reason: it seems possible *in each of these senses* for rational action to be produced without deploying irreducibly first-person indexical thoughts. If so, some IAR's are complete. To show how this might happen, they offer a model of action production, which they call the *Action Inventory Model*.

> **The Action Inventory Model**: Every agent has a very wide range of third-person beliefs and desires that give rise to third-person intentions, which in turn rationalize or motivate actions (via their recognition). Not all of these intentions are going to produce action, at least in normal cases (perhaps in a god they would). This is because a given agent has an "action inventory": a range of actions that he can perform. An agent constantly seeks to match his intentions with his action inventory, and when he finds a match, action occurs. When there's no match, the intention idles, and doesn't motivate or rationalize action. So the Selection Problem is solved by appealing to the physical or psychological constraints of the agent: only certain actions result because only certain actions were available in the first place." (p. Cappelen and Dever 2013, p. 50).

On this account, acting simply consists in matching third-person intentions with available actions in one's action inventory. But neither one's action inventory, nor one's third-person intentions, need consist of irreducibly first-person indexical thoughts. Because it is possible that action works in this manner, (IIC) fails. Perry's argument thus fails to establish anything about thought, much less the

[59] E.g., Bermudez (2017) defends (IIC) explicitly, and McGinn (1983) contends that "indexical concepts are ineliminable because without them agency would be impossible: when I imagine myself divested of indexical thoughts, employing only centreless mental representations, I *eo ipso* imagine myself deprived of the power to act" (McGinn 1983, p. 104).

irreducibility of first-person thought. The famous examples, which feel so intuitively gripping, are merely Frege cases that support (GO).

4.2.3 A Reply: Clarifying Perry's Argument.

Cappelen and Dever's criticism of Perry fails, but showing why requires clarifying Perry's argument. I grant that the argument involves a Frege case,[60] but deny that things end there. Perry offers a two-step argument, and the charge that he merely produces a Frege case misses the second step. The first step of Perry's argument involves identifying an intuitive change in belief, which he describes by noting a failure of substitutivity, and by using the language of 'essential indexicals." The second step traces the "essential indexicals" used when describing the change in belief back to the thoughts themselves. To do so, the argument deploys a contingent, but explanatory, connection between thought and the language that agents canonically use when expressing thought.

To show that Perry's argument is a mere Frege case, Cappelen and Dever argue for the (GO)-reading of cases like *Sloppy Shopper*. To illustrate what more Perry's argument establishes, and how, recall the considerations adduced in favor of the (GO)-reading.

First, Cappelen and Dever argue that because cases like *Sloppy Shopper* and *Superman/Clark Kent* are structurally identical, each (partially) establishes (GO). But proponents of Perry's argument have a ready response: the examples are relevantly different, as the change of belief in *Sloppy Shopper* plays a different role in rationalizing Perry's action than does the change of belief in *Superman/Clark Kent*. This can be seen by drawing out contrasts in which actions get rationalized. In *Sloppy Shopper*, the change of belief rationalizes Perry's acting (as opposed to not acting), whereas the change of belief in *Superman/Clark Kent* rationalizes the agent telling Superman to clean up after himself (as opposed to telling Clark Kent). But when we ask about what rationalizes the agent telling Superman to clean up (as opposed to not acting), the explanation involves irreducibly first-person thoughts for the reasons made famous by Perry's examples. Consequently, there is a relevant difference between the cases, and both cases support or assume (IO).

[60] Or at least involves a failure of substitutivity which can be represented as a Frege case. For Perry (1979), the failure of substitutability of co-referential terms isn't primarily meant to establish an opaque context per se, but rather to give us some way of getting a hold on key features of the change in belief that occurs in cases like *Sloppy Shopper*.

Cappelen and Dever should reject this reply on the basis of their argument against any necessary connection between action explanation and irreducibly first-person thoughts. However, this shows that Cappelen and Dever's case for the GO-reading really hinges upon the success of their second consideration.

Second, Cappelen and Dever argue that there are no suitable inference principles leading from the explanation of action to irreducibly first-person thought, as there are no philosophically deep or interesting connections between the two. This claim is mistaken, but not because some version of (IIC) is true. Assuming that Perry's argument is deductive in form, and turns on more than "shallow contingencies," they set out to find principles which capture the necessary connection between first-person thought and action. No such principles are to be found, because it is possible for action explanation to proceed without irreducibly first-person thought. Whether the Action Inventory Model offers an adequate explanation of human actions is contested,[61] but for the purposes of rejecting (IIC), all that needs to be shown is that it isn't necessary (in the operative sense) that action explanation requires essentially indexical thought.

Even setting aside the Action Inventory Model, this seems simple enough to do. We can imagine a being whose capacity for irreducibly first-person thought can be selectively disabled, even while the being continues to act. If so, then at least some of the being's actions can be explained without irreducibly first-person thoughts. Alternatively, we can imagine building a robot that performed actions, but wasn't capable of having irreducibly first-person thoughts at all.[62] Here the role typically played by irreducibly first-person thought is accomplished instead by the connections hard-wired into the robot's architecture. In either case, it seems possible to explain rational action without irreducibly first-person thought, which is enough to reject the relevant readings of (IIC).

While there are lingering issues about whether it is possible to explain *genuinely* rational action in this manner, these seem beside the point. Or at least beside Perry's point. Of cases like *Sloppy Shopper*, Perry says "my change in beliefs seems to explain my change in behavior. My aim here is to make a key point about characterizing this change, and of beliefs in general" (Perry 1979, p. 3). So, while he often characterizes what is distinctive about the change in belief in linguistic-sounding terms like "essential

[61] See Bermudez (2017), Ninan (2016), and Torre (2018) for critical discussion.
[62] Teller (2011) pursues this thought.

indexical," the primary focus of Perry's essay is not on understanding essentially indexicality in attitude ascriptions, but on understanding a change in belief itself.[63] Perry reiterates this point when explaining what he takes himself to have shown: "we use sentences with indexicals or relativized propositions to individuate belief states, for the purposes of classifying believes in ways useful for explanation and prediction. That is, belief states individuated in this way enter into our commonsense theory about human behavior and more sophisticated theories emerging from it…A good philosophy of mind should explain this in detail; my aim is merely to get clear about what it is that needs explaining" (*ibid.*, p. 18). As I understand it, Perry's account of 'what needs explaining' involves high-level empirical claims about which belief states explain which actions, and thus which belief states the agent is in. These generalizations needn't establish necessary connections, as psychological generalizations standardly have exceptions and are formulated using ceteris paribus clauses. But this doesn't stop them from being explanatory. Folk psychology is a body of theory which, if vindicated, is generally taken to be capable of predicting, understanding and *explaining* behavior. As such, we have located a set of candidate principles with which to run Perry-style arguments for irreducibly first-person thought: the contingent, but nonetheless explanatory, connections found in commonsense psychology.

If such a principle can be articulated and defended, then Cappelen and Dever's case for the (GO)-reading fails. Which kinds of principles might drive the second step of Perry's argument? Consider the high-level psychological generalizations which connect the language that agents canonically use when expressing their belief state with the way in which they think about themselves when self-attributing that state. The precise form that these principles take will likely depend upon the correct account of irreducibly first-person thought. Moreover, for the purposes of replying to Cappelen and Dever, I'm more interested in defending the general empirical approach to Perry's argument than in defending any specific principle. But, to give an example, consider the account of self-attached knowledge that Perry elaborates in later work. Famous examples notwithstanding, Perry tells us that "ordinarily all one's knowledge about oneself is integrated around a special sort of idea or notion of

[63] We shouldn't collapse Perry's argument for irreducibly first-person thoughts into Perry's argument against the doctrine of propositions. Much of the discussion of "essential indexicality" in Perry 1979 is due to his choice of target: the doctrine of propositions. According to the doctrine of propositions there is an isomorphism between the structure of that clauses, and the structure of beliefs as beliefs are individuated by their objects, and that the objects of beliefs are picked out by that-clauses. Thus, Perry's argument against the doctrine might well be read as focusing upon the role of "essential indexicals" in belief ascriptions. But the positive account that Perry offers distinguishes beliefs from belief states, and arguing that "there is not an identity, or even an isomorphic correspondence, but only a systematic relationship between the belief states one is in and what one thereby believes" (Perry 1979, p. 40).

oneself that we express with "I"," (Perry 1998, p. 333). This integration is achieved when ideas related to the self are associated with a self-notion, and self-attached knowledge is the kind of first-person knowledge which is canonically expressed with the word 'I'.

This suggests the following principle:

> **(P)** If articulate, sincere agents express a belief by using first-person "essential indexicals," then the agent's belief state deploys a self-notion.

Perry's argument proceeds as follows. The first step delivers not only the claim that action explanation contexts are opaque, but the claim that the belief which explains Perry's action is expressed with an "essential indexical." The second step deploys (P) to take us from the language that is canonically used to express belief, to a claim about the nature of the belief state itself. The argument then concludes by explicitly relating belief states which deploy a self-notion to irreducibly first-person thought.

Instead of evaluating the empirical support for (P), I will make two preliminary points. First, we know roughly how to defend a principle like (P): by looking at candidate cognitive architectures for self-knowledge. If the belief states expressed by first-person essential indexicals are instances of self-knowledge, then to the extent that good models of self-knowledge have the integrated structure that Perry proposes, (P) is consistent with, or even justified by, available empirical evidence.[64]

Second, the standards for evaluating empirical models of both cognition and action are different from those employed by Cappelen and Dever in their discussion of Perry. For example, models of action production are not seeking to explain how action *must* be produced, but how to *best* explain human action production. Appreciating this fact clarifies the status of the Action Inventory Model in debates about irreducibly first-person thought. The main problem with the model isn't that action couldn't operate in the way it suggests, but that the model fares poorly as an empirical account of action production. It is underdeveloped, computationally inefficient, and fragile. It needs to be further developed to capture basic features of rational action, e.g. the fact that we don't simultaneously perform incompatible actions. More problematically, storing and exhaustively checking the database

[64] Horgan and Nichols (2016) argue that irreducibly first-person thought provide computationally efficient accounts of episodic memory and planning. While their discussion doesn't directly advocate for (P), it is suggestive of the kind of empirical support that might be developed.

of the Action Inventory Model is computationally inefficient, and if anything goes wrong in reading intentions or available actions, then no match will occur. Thus the model is inefficient and fragile. This isn't yet a defense of the Perry-style argument for irreducibly first-person thought, but it helps clarify why it is that proponents of such arguments remain unimpressed by the Action Inventory Model.

In short, Perry's argument is a broadly empirical argument, and whether or not it is compelling will ultimately turn on facts about how first-person thought is implemented. Cappelen and Dever briefly consider this response before dismissing it as hopeless. Perusing the existing literature, they deem Perry's argument to be a non-empirical argument, as no empirical work is being cited by proponents of the argument. Further, they claim that there is no " 'doing psychology about very general truths about human beings' without empirical backing" (Cappelen and Dever 2013, p.55). As such, they summarily dismiss principles of inference that establish merely contingent connections of the kind that might be empirically discoverable.

It is worth noting that I'm articulating Perry's argument itself, and not a concession that proponents of the argument might make. For some subject matters, an empirical argument is the right kind of argument to give, and proponents of irreducibly first-person thought owe no apology for the broadly empirical nature of Perry's argument. More importantly, Cappelen and Dever's criticism fails: they have rejected *empirical* arguments for irreducibly first-person thought on the grounds that previous proponents of the argument haven't adequately supported a key premise. But, an empirically-based argument might well succeed with different support. Moreover, philosophers have long argued for general truths about human psychology without explicitly citing empirical support. On the basis of (non-empirical) epistemic desiderata, philosophers can, have, and will continue to argue for certain accounts of perception or cognition. Further, the character of Perry's argument is familiar, naturally located in the tradition of "speculative psychology" that Jerry Fodor advocated in the introduction to *The Language of Thought*. While there are many ways in which such an inquiry can go wrong, we can't reject it out of hand either.

Cappelen and Dever argue that not only are there no irreducibly first-person thoughts, but that "the entire topic is an illusion—there's nothing there" (Cappelen & Dever 2013, p. 3). Following Perry, I've argued that "essential indexicality" is a broadly empirical topic in the philosophy of mind. Further,

I've gestured at how to further develop arguments for irreducibly first-person thought. Such thought is not postulated merely to solve "the problem of essentially indexical thought," as articulating a *distinctive* problem require making stronger assumptions about the nature of propositions, attitudes or content, than are needed to state Perry's general argument.[65] While such an argument might ultimately want further empirical backing, this suggests only that the case for irreducibly first-person thought is currently incomplete, and not any flaw with the argument itself.

All told, Cappelen and Dever's criticism offers a useful reminder of the broadly empirical character of Perry's argument. While the first step of Perry's argument can be represented as a Frege case, the second step deploys a principle that captures a contingent, but philosophically interesting, connection between first-person thought and action. Thus, Perry's famous example provides the first half of a two-step argument, the second step of which is no mere Frege case.

Of course, there are any number of other reasons that a philosopher might give by way of rejecting the arguments from the CPL tradition. But I don't find them any more compelling than the one examined here. Until one of them succeeds in establishing a serious problem, we lack compelling reasons to reject the kinds of arguments for irreducibly first-person thought given by the CPL tradition.[66] Whether or not irreducibly first-person thought itself is ultimately defensible, the arguments for it are not themselves objectionable.

4.3 Is there a problem with Irreducibly First-Person Thought itself?

A second type of objection claims that there is a problem with irreducibly first-person thought itself. That is, whether or not the arguments given in favor of irreducibly first-person thought are problematic, such thought is itself problematic and must be rejected. Thus, independently of the success of arguments for irreducibly first-person thought, there are reasons to reject it.

One such argument contends that if irreducibly first-person thought has a distinctive kind of content, then such thoughts cannot be communicated. But we readily communicate such thoughts, thus, irreducibly first-person thought is a fundamentally confused notion that ought to be rejected. My

[65] Accordingly, showing that there is no distinctive problem of essentially indexical thought isn't a mark against Perry's argument for irreducibly first-person thought.
[66] Nor, by extension, should skepticism about these arguments for the irreducibility of first-person thought hamper my extension of them to argue for the irreducibility of second-person thought.

arguments for irreducibly first-person thought don't assume that the mode of presentation is somehow or another built into the content. Instead, I've been using a more neutral notion of irreducibly first- and second-person thought. But I haven't given us any reasons to reject such an account either, so in this section, I'll be defending a version of the content reading of irreducibly first-person thought in order to defend the generality of the *Basic Argument*. This section develops a communication-based objection and explains why it ought to be rejected.

4.3.1 The Puzzle of Communicating Essentially Indexical Thought.

Suppose for the purposes of argument, that *essentially indexical thoughts* have distinctive contents, which I will call *essentially indexical contents*.[67] Any theory of thought content should meet (at least) two desiderata. Thoughts explain action, so a good theory must yield contents "rich" enough to allow us to explain an agent's actions in terms of their thoughts. Also, thoughts can also be expressed linguistically, so a good theory must yield contents "lean" enough to be communicated linguistically by a speaker.

Essentially indexical thought presents a challenge to theories of thought content, as these desiderata pull apart. Consider a paradigmatic essentially indexical thought, e.g., the thought expressed by John Perry's sloppy shopper upon uttering "I am making a mess."[68] The content of this self-locating thought (putatively) captures the first-person perspective of the thinker. While Perry himself didn't think that essentially indexical thoughts have essentially indexical contents, one might. For example, David Lewis argues that these thoughts should be modeled using centered possible worlds to capture the thinker's "perspective."[69]

Lewis's account also illustrates a problem for accounts of essentially indexical thought—essentially indexical thoughts seem to be readily communicated via language, but their centered content is not

[67] This supposition is merely for the purposes of this section. Generally speaking, I employ the broader notion of irreducibly first-person thought.
[68] Perry (1979)
[69] Lewis (1979). Throughout the paper, I model the content of essentially indexical thought using centered possible worlds. I do so not to endorse Lewis's account, but to illustrate the issues that communication poses for accounts of essentially indexical thought. That said, I do think that various accounts of essentially indexical thought can agree that having *centered content* (e.g., content that can be fruitfully modeled by what Egan 2007 called *interesting* centered-world propositions) is a minimal condition distinguishing essentially indexical content from other kinds of content.

what is communicated by the shopper's utterance "I am making a mess." These conflicting desiderata yield the *Puzzle of Communicating Essentially Indexical Thought*.

> **(4)** Essentially indexical thought has distinctively centered content.
> **(5)** Essentially indexical thought can be communicated linguistically.
> **(6)** Successful communication involves transmitting a single belief content from speaker to addressee.
> **(7)** What hearers understand lacks distinctively centered content.

(4) is a widely endorsed claim about essentially indexical thought. Consider a paradigmatic kind of essentially indexical thought: *de se* thought. *De se* thoughts are thoughts about oneself in which one represents oneself as oneself. Examples include Ernst Mach's realization that he is the shabby pedagogue that he sees in the mirror, and David Kaplan's realization that his pants are on fire.

Many philosophers think that essentially indexical thoughts have distinctive content, *essentially indexical content*, that captures the thinker's perspective. For example, John Perry and David Lewis have famously argued that *de se* thoughts have contents that are "self-locating" in that they are about who, when and where the thinker is.[70] For current purposes, I will model the content of essentially indexical thought using centered possible worlds, or *centered contents*, which lack globality and accessibility.[71]

More generally, essentially indexical contents have two distinguishing features. To introduce them, I assume a framework of centered possible worlds in which propositions are true or false at points of evaluation. A point of evaluation is an ordered pair of a world and location within that world $<W_n,L_n>$. Locations can be understood as persons, objects or places (which may or may not be temporally indexed). The content of an essentially indexical thought is not true at every point of evaluation within a world, thus it lacks globality.

> *Globality*: A propositional P is global in world W only if P has the same truth value at any location ($L_1 \ldots L_n$) in W.

Further, if we say that a proposition that is accessible by any pair of agents with identical knowledge* and concepts* has the property of accessibility, then propositions with essentially indexical content are not accessible.

---

[70] Of course, different theories of content "capture the thinker's perspective" in different ways. The theories of Perry and Lewis illustrate two different ways that one might seek to capture the "self-locating" aspect of self-locating thought.
[71] Globality and Acessibility are defined in Chapter 3.

> *Accessibility*: If a subject S_1 and S_2 in W share all the relevant concepts* and knowledge*, and S_1 believes proposition p, then S_2 can also believe p.[72]

Combined, a content C is an essentially indexical content iff a proposition with content C lacks globality and accessibility. Different theories of content (using possible worlds, structured propositions, etc.) might meet these two conditions. For current purposes, I will model the content of essentially indexical thought using centered possible worlds, or *centered contents*, which lack globality and accessibility.[73]

(5) is also widely accepted. One can express essentially indexical thoughts using indexicals,[74] and addressees can readily understand the thoughts so expressed. We routinely express *de se* thoughts by using 'I' or 'he*'. It is utterly unremarkable when people understand what we intended them to by making such utterances. In this *minimal sense*, essentially indexical thought can be communicated linguistically.

What does it mean to say that essentially indexical thought can be communicated by a speaker to an addressee? (6) derives from the *Naive Account of Communication*. According to this account, communication succeeds iff there is a single content C that: (i) begins as the content of the speaker's belief, (ii) gets transmitted as the content of the speaker's utterance by which she expresses this belief, (iii) is understood by the addressee, and (iv) the addressee forms the belief that the speaker believes that C. Thus, on the Naive Account, successful communication involves a single content C that makes the journey from speaker to addressee.[75]

---

[72] The '*' places a restriction on the kind of knowledge and concepts in question. Knowledge* is knowledge of globally true propositions, and concepts* are the kinds of concepts that might be constituents of globally true propositions.

[73] Footnote 2 clarifies my (lack of) commitment to modeling the content of essentially indexical thought with centered possible worlds. Otherwise, the connection between essentially indexical content and centered contents should be fairly clear: Globality is a feature of propositions. Propositions that have Globality have uncentered contents, and propositions which lack Globality have centered contents.

[74] That one can express essentially indexical thoughts is widely, but not universally, accepted. For example, Gottlob Frege and David Lewis are often read as claiming that you cannot communicate such thoughts. I won't argue against this position here. Instead, in the remainder of the paper I assume that essentially indexical thoughts can be communicated in either the minimal or stringent senses that I introduce shortly.

[75] It is worth noting that the *Naive Account* makes claims about both what gets communicated, and what counts as successful communication. Although I end up rejecting the naive model, I agree that any good account of communication should address both of these.

Endorsing (6) differentiates distinct readings of (5). The *Minimal Reading of (5)* corresponds to the minimal sense that I stated previously (when we make utterances that express essentially indexical thoughts people understand what we intended them to). (6) replaces this common-sense notion of communication, yielding the *Stringent Reading of (5)*, which claims that essentially indexical thought can be communicated linguistically in the sense of communication specified by the Naive Account.

(7) points out something about what gets communicated, regardless of one's reading of (5): essentially indexical contents aren't what an addressee understands when a speaker expresses an essentially indexical thought. For example, suppose that John Perry fixes his torn sack of sugar, chuckles, and says to the store manager "I was the one making a mess all along!" The manager will have no trouble understanding Perry, and won't self-ascribe the property MAKING A MESS. Perry's realization that *he himself* was the one making a mess seemingly captures his self-locating realization, but when Perry communicates this thought by using the first-person indexical, the content communicated isn't this distinctively centered (viz., self-locating) content.

4.3.2 Decentering Responses.

How should one resolve this puzzle? Let's assume (5), in some sense, and (7). At issue, is whether to reject (4) or (7), and why. *Decentering Responses* reject (4), denying that "essentially indexical thoughts" have centered content.[76]

For example, the *Basic Kaplanian Response* (BK) rejects (4) on the grounds that the content of essentially indexical thought is not interestingly different from other forms of thought.[77] This conclusion is reached by recruiting basic distinctions from Kaplanian semantics.[78] (BK) begins by observing that some language is context-sensitive. When Jane utters the sentence "I need coffee", 'I' refers to the speaker of the utterance (Jane). When John uses the same sentence, 'I' still refers to the speaker of the utterance, but now the speaker is John. The variability of the reference of 'I' stands in contrast to the stability of the reference of names (like 'Jane'). Kaplanian semantics captures this difference by distinguishing between the character of a sentence and the content of a sentence in a given context.

[76] Strictly speaking, because essentially indexical thoughts have essentially indexical contents, decentering responses deny that there are essentially indexical thoughts. They still make claims about the content of so-called "essentially indexical thoughts," so I continue to use the term in this section. Nothing substantive hangs on it.
[77] As I read them, Cappelen and Dever (2013) and Magidor (2015) each endorse (BK).
[78] As introduced in Kaplan (1989).

The character of a sentence is a function from context to content, while the content of a sentence (in a given context) determines that truth value of that sentence (in a given context) in a possible world. Similar claims apply to parts of sentences. The character of 'I' is variable (yields different referents in different contexts of use), while the character of 'Jane' is stable (yields the same individual across contexts of use). But Jane uttering 'I need coffee' or 'Jane needs coffee' both express the same content: the individual Jane, and a property NEED COFFEE.[79]

Applying these distinctions to both thoughts and sentences, (BK) undercuts (4) by denying that essentially indexical thought is distinctive. Let the analog of the utterance of a sentence be the thinking of a thought. By distinguishing content and character, (BK) denies that the content of an essentially indexical thought in context is an essentially indexical content, while allowing that the character of some thoughts is variable (their reference shifts from context to context). The resulting account of the content of thought and language accepts (5)-(7), thus denying (4).

Two main problems for BK involve adequacy and arbitrariness. The adequacy problem is that (BK) requires that one specify the character of sentence/thought constituents that are fixed by convention and don't vary by context of occurrence. This is non-trivial as paradigmatic instances of linguistic indexicals like 'I' and 'that' have proven problematic in this regard,[80] and because otherwise promising accounts don't work for thought (e.g., the character of 'you' is "the addressee of this utterance").

The arbitrariness problem contends that when communicating a thought, distinguishing the character and content of the thought is arbitrary. Kaplanian semantics doesn't generalize to thought for free, and the mental analogs of utterance, context of utterance and what is said aren't straightforward. Further, there are reasons to think that such an account can't be developed. For example, in Perry's sloppy shopper example, the analog of both content and character play a role in explaining Perry's behavior. In Lewis's *Two Gods* example, the analog of both content and character seem to be something that the God learned. Absent reason to think otherwise, the intuitions driving the existing literature

[79] The exact details concerning the propositional constituents aren't important here. Kaplan himself endorses an account of singular propositions, but one needn't do so to endorse (BK).

[80] Much of the contemporary discussion of indexicals in the philosophy of language is focused on the viability of Kaplanian accounts. (For examples of problems, see Nunberg 1993 on descriptive indexicals or Cohen and Michaelson (2013) on the Answering Machine Paradox).

on essentially indexical thought suggest that the content/character distinction is arbitrary when applied to thought.

There are other decentering responses that I haven't discussed.[81] But my main aim is to demotivate de-centering responses, not definitively reject them. By endorsing (5)-(7), decentering accounts run the *Puzzle* as an argument for the rejection of (4). If there is an alternative explanation of the communication of essentially indexical thoughts that rejects (7), it undercuts this argument.

4.3.3 Two Ways to Communicate Essentially Indexical Thought*.

The remainder of this section gestures at how one might construct this explanation. (6) should be rejected independently of any resolution to the Puzzle, because the Naïve Model offers too restrictive an account of communication. Non-assertoric uses of speech often succeed communicatively without transferring a content C to the addressee (e.g., "Would you pass the salt?"). Even restricting ourselves to assertion, communication occurs in the presence of insincere speakers and skeptical addressees, who block the transmission of a single content C from speaker to hearer. Putting these "messy cases" aside, I'll argue that assertions expressing essentially indexical thought can succeed communicatively without transferring a content C from speaker to addressee.

Doing so requires a different account of successful communication. Call any account which allows that centered contents play some role in the communication of essentially indexical thoughts a *Centering Response*. Two varieties of centering response are Surrogate Responses and Re-centering Responses.[82] They differ in the role that centered contents play in the successful communication of essentially indexical thoughts, and I argue that successful communication involves a form of re-centering.

Surrogate Responses (SR) maintain the spirit of the Naïve Model by claiming that in successful communication, a single content is conveyed from speaker to hearer. (SR) adds that when the speaker's utterance cannot convey the literal content of her essentially indexical thought (which might be unshareable), it conveys a surrogate content sufficiently similar to the content of the speaker's thought to count as communicating her thought. Thus, (SR) endorses:

[81] Most prominently Robert Stalanker's account, which aims to show why "the distinctive character of self-locating attitudes does not imply that there is a distinctive kind of self-locating content" (Stalnaker 2016, p. 122). A full-treatment version of this section would discuss these other de-centering responses.
[82] Recanati (2016) uses the framing and terminology used in this section.

> **(SR-P)** A speaker P successfully communicates her thought T with content C iff P's assertion expressing T transmits C (or some suitable surrogate S*) from P to the addressee.[83]

Proponents of (SR) elaborate the conditions for suitable surrogates.[84] Importantly, S* can be a centered content, so long as it a shareable centered content. For example, if Perry utters "I am making a mess" the content transferred might be the property of being party to an utterance U in which the speaker of U self-ascribes the property MAKING A MESS. Because both speaker and addressee are party to the utterance, this centered content can be shared.[85]

Re-centering Responses (RR) more thoroughly reject the Naïve Model.[86] (RR) contends that when expressing an essentially indexical thought, the speaker's utterance faithfully expresses its centered content C. Upon hearing and accepting the utterance, the addressee comes to believe a different content R*, which is suitably related to C. The "suitably related" content R* is created by transforming C into R* by using the addressee's relation to the speaker to re-center C. Communication succeeds when the addressee comes to believe R* on the basis of the speaker's utterance.

Setting aside the issue of how re-centering takes place,[87] a content R* is *suitably related* to C when it meets (at least) two conditions. Communication involves not only transmitting facts, but getting things done. Skilled communicators "want the hearer not only to think about the right object, but to think about it in the right way, a way that connects with what the speaker wants the hearer to do."[88] Thus, R* must refer to the same object(s) as C, and R* must get the hearer to think about its object(s) "in the right way."[89]

---

[83] Put alternatively, one successfully communicates a thought with content C by meeting the conditions laid down by the Naive Model for either content C or S*. This is an expansion of the Naive Model, which made claims about (i) what is communicated, and (ii) what successful communication consists in. (SR) relaxes (i) to include the surrogate S*.

[84] E.g., Kolbel (2013) and Torre (2010).

[85] Accounts vary, but the general strategy of (SR) is clear enough—when the contexts of the hearer and addressee are relevantly different, centered contents that reflect context aren't communicable. The surrogate is such that it places the speaker and hearer in the same relevant context (rendering the relevant centered contents communicable).

[86] E.g., Weber (2013), Recanati (2016).

[87] There are various ways in which this process of "transform-and-recentre" might be accomplished. See Recanati (2016) for discussion.

[88] Perry (2011). Perry points out that skillful communicators understand the *cognitive path* that leads from the speaker's conception of the object to the hearer's conception (p. 64).

[89] Here I'm not defending a general account of what "the right way" is here, because the argument that follows doesn't require one. But one need not give such a general account to give substance to the "right way" either. For example, Perry (2011) understands such assertions as taking place in *information games*. What counts as the "right way" might then be relative to the information game being played.

I prefer the latter option, but for the purposes of this chapter, there is no need to pick sides. Either way, there is nothing in the puzzle that suggests that we cannot communicate essentially indexical thoughts, or that they impose any special burden upon those who endorse them. There is research to be done in explaining communication, but this is hardly an objection to essentially indexical thought.

The problems posed by other objections to the irreducibility of first-person thought aren't any more pressing. Most continue to focus varieties of irreducibly first-person thought that locate the irreducibly first-person mode of presentation in the content of the thought, e.g. "essentially indexical thought" (as I've used the term in section 3). Notably, Ruth Millikan has argued that "so-called "essential indexicals" in thought are indeed essential, but they are not indexical" (Millikan 1990, p. 723). But in the context of the permissive manner in which I'm using the phrase "mode of presentation," Millikan's interesting arguments look more like a vote for a certain kind of irreducibly first-person thought than an objection to a premise of the *Basic Argument* itself.

4.4 Conclusion: Towards an Account of Irreducibly Second-Person Thought?

While this chapter has offered only a selective summary of the "first-person skepticism" which might lead one to question premise (1) of the basic argument, I've organized some of the main causes of second-person skepticism, and responded to two of the more prominent varieties. Doing so illustrates both the kinds of objections that philosophers have made against the arguments for irreducibly first-person thought, and the argumentative resources that its proponents might utilize in responding. I conclude that there are neither devastating objections to CPL-style arguments for irreducibility, nor is there reason to think that irreducibly first-person thought itself is inherently problematic. This assessment is meant to be compatible with both the claim that the arguments of the CPL tradition often stand in need of clarification, and the claim that further work is needed to understand the role of irreducibly first-person thought in explaining action and communication. More generally, though, the discussion of this chapter is meant to bolster a philosophical orthodoxy which already accepts the irreducibility of first-person thought.

What bearing does this have on the *Basic Argument*? In addition to defending premise (2) of the argument, and I've now rebutted the more prominent objections to (1) as well. I conclude that the *Basic Argument* succeeds. This is not to claim that all philosophers are now rationally compelled to

endorse irreducibly second-person thought. Rather, it is to say that the *Basic Argument* takes the often endorsed claims of the irreducibility of first-person thought, and uses commonly given reasons for endorsing irreducibility to support to a novel conclusion: the irreducibility of second-person thought. Further, the *Basic Argument* does this without compromising its broad appeal. Nothing hinges upon contentious claims about content, the relation between language and thought, etc. Thus, so far as the conclusion of the *Basic Argument* is concerned, philosophers might fill in the details in a variety of ways.

However, this sort of neutrality begets a lingering worry for some philosophers: I've argued for irreducibly second-person thought without saying exactly what it is! While this is advantageous in certain respects, it also suggests that the task of defending irreducibly second-person thought is incomplete. The matter is pressing, as the implications of irreducibly second-person thought might hinge upon the details of such an account. What, broadly speaking, are the available options?

While all forms of irreducibly second-person thought involve an irreducibly second-person mode of presentation, an account of irreducibly second-person mode of presentation might take various forms. These largely hinge upon one's chosen account of mode of presentation. One might take the irreducibly second-person mode of presentation to be a part of the thought's content. Alternatively, one might take a mode of presentation to be something else (ways of believing a proposition, functionally characterized vehicles of thought, etc.) that although related to the content of a thought, is not, strictly speaking, a part of it.

This gives the appearance of a looming choice for proponents of irreducibly second-person thought: is the distinctively second-personal mode of presentation a part of the content of irreducibly second-person thought, or a part of something else? So far as accepting the *Basic Argument* is concerned, one might come down in either direction. But different accounts of irreducibly second-person thought enjoy corresponding strengths, while facing corresponding challenges. The next chapter examines one way of proceeding: by treating the second-person mode of presentation as a mental file. Although this is certainly not the only account available, it has some appealing features, particularly for those philosophers who are a bit uneasy about modes of presentation in the first place.

Chapter 5: The Second-Person Mode of Presentation as Mental File

Frege taught us that cognitive significance outstrips reference. The cognitive significance of a thought cannot be captured merely by the referent of the thought, but by the way a thought presents this object—its mode of presentation. However, since Frege introduced modes of presentation, their nature and necessity have remained contentious topics, and the available theoretical options for understanding them are both complicated and numerous. The most deflationary treatments simply deny the need for modes of presentation altogether, squashing the issues raised by Frege into an account of the language that we use to ascribe attitudes to thinkers. More satisfactory accounts acknowledge the need for modes of presentation, but, having done so, treat them in various ways. Modes of presentation might be placed in the propositional content itself (e.g., as Fregean senses, or more loosely speaking, as centered contents). Alternatively, one might take the propositional content to be exhausted by the referent itself, and instead postulate "ways of apprehending" this content (e.g., as Neo-Russellians have done), or one might differentiate a mental state type from its content, and claim that the mode of presentation plays a role in individuating mental states, but needn't show up in the content of a state. Even further, one might deploy a pluralistic notion of content, and claim that the mode of presentation plays a role in one kind of content broadly construed, but not in the truth conditions of the thought narrowly speaking.

These options might not be exclusive, and this list clearly isn't exhaustive. Fortunately, the *Basic Argument* doesn't turn on the details of any specific way of treating modes of presentation. Thus, we needn't to sort out issues concerning what modes of presentation *are* in order to appreciate the argument for the irreducibility of second-person thought. But this doesn't mean that an account of modes of presentation isn't eventually needed. And in particular, it doesn't mean that defending an account of the irreducibly second-person mode of presentation doesn't yet have a role to play in dispelling a certain kind of skepticism about irreducibly second-person thought.

To see why, consider two ways in which one might interpret the irreducibly second-person mode of presentation. One might take it to be a part of the thought's content. For my purposes, this is the

easier of the two cases, as it is easier to see what irreducibly second-person thought is, should there be any. If one is generally inclined towards capturing modes of presentation in the content of a mental state, and follows the *Basic Argument* to its conclusion, then one simply needs an account of the distinctively second-personal contents of irreducibly second-person thoughts. While this isn't a trivial task, there is no reason to think that the second-person mode of presentation, as opposed to other modes of presentation, poses a particular challenge.

Should one deny that modes of presentation are a part of the content of a thought, the situation is a bit more complex. Given the permissive way that I am using the term "mode of presentation," this denial shouldn't yield any reason to be skeptical about the *Basic Argument*. Whatever modes of presentation turn out to be (ways of believing a proposition, functionally characterized vehicles of thought, etc.), so long as there can be irreducibly second-person modes of presentation, then one ought to be able to understand irreducibly second-person thoughts in one's chosen terms. Nonetheless, philosophers who are generally leery of modes of presentation might also be skeptical about second-person thought. By operating with such a permissive understanding of mode of presentation, it might seem like the *Basic Argument* has simply employed a potentially problematic notion without saying what it is. For philosophers who are skeptical about modes of presentation, the broad appeal of the *Basic Argument* will look more like a liability.

Before signing on for anything called irreducibly second-person thought, this latter kind of philosopher will want to know more about what modes of presentation are, and, in particular, what irreducibly second-person modes of presentation might be. This request is reasonable, and this chapter provides the beginnings of a reply. Although I won't argue that philosophers must endorse the account that I give here,[90] this chapter will sketch an account of the second-person mode of presentation that is interesting, plausible, and aimed at satisfying those whose skepticism about irreducibly second-person thought stems from their skepticism about modes of presentation.

I earlier argued that the second-person mode of presentation represents its object as *an agent with whom I am interacting*. In this chapter, I'll interpret this gloss as a non-descriptivist mode of presentation, and

[90] That is to say, I won't be arguing against alternative ways of interpreting the second-person mode of presentation. For example, I won't be arguing against the first of the two views just introduced, which claims that one should keep modes of presentation out of the content of a thought in the first place. Nor will I argue against other "non-content" ways of understanding modes of presentation.

further assume that non-descriptivist modes of presentation are mental files. In particular, I'll appeal to Francois Recanati's version of the mental files framework[91] to articulate an account of a distinctive type of file which I contend captures the second-person mode of presentation. Mental files are metaphorical "containers" for information about specific objects. Thus, we might distinguish three aspects of a mental file: the container itself (the file type and token), the content (body of information), and the referent of the file. On Recanati's account, file types are (roughly) individuated by an epistemically-rewarding (ER) relation that the subject of the thought stands in to the object of the thought. This ER-relation populates the file with information that is presupposed to be about a unique object. But the referent of the file is not determined by satisfying the body of information in the file. Instead, the referent of the file is the object that stands in the relevant ER-relation to the thinker. I'll argue that in the context of reciprocal interactions, thinkers can stand in a relation of *direct engagement* to their interlocutors. This relation affords the capacity for a distinctive kind of file, an *interlocutor file*, which captures the cognitive significance of irreducibly second-person thought. I call the resulting account the *second-person mode of presentation as interlocutor file account* (or SPIF account).

The chapter proceeds as follows. In *section 1*, I will introduce Recanati's basic framework for understanding mental files as the vehicle of singular thoughts. Section 2 will distinguish two kinds of mental files, object files and self files, and the distinctive kinds of epistemically-rewarding relations upon which each is based. *Section 3* identifies a further kind of epistemically-rewarding relation, direct engagement, that holds between agents undergoing certain kinds of interactions. Using this relation, I articulate an account of interlocutor files, and argue that the second-person mode of presentation is identical to such files. Having developed the SPIF account, *section 4* will examine some criticisms of it. I conclude with some reflections on the epistemic implications of the account. Although the SPIF account isn't the only way to understanding irreducibly second-person thought, it is an appealing one, and it brings with it some potentially interesting new lines of research.

5.1 Mental Files as the Vehicles of Singular Thought.

Singular thoughts are thoughts that are intuitively about a particular object, as opposed to thoughts about whatever object happens to satisfy the descriptive content of the thought. *Descriptivism* contends that a thought refers to an object when it represent properties that are instantiated by the object, and

[91] Recanati 2012, 2016.

thus refer to it.[92] Thus, according to descriptivism, a thought refers to a unique object because an object uniquely instantiates some property (or properties) that the thought represents as being instantiated. Paradigmatically, these reference-fixing properties can be expressed using a definite description, e.g., 'the tallest man in Australia'. Such accounts face a challenge: it isn't clear that such descriptions can capture the singularity of thought without employing terms that refer directly to elements of the very scene one is in (e.g., 'actual', 'that', 'in front of me', etc.). To use an example from Quine, if I desire a specific sloop, it seems like my desire is for that very sloop itself, and not merely for "relief from slooplessness," which might be met by acquiring any sloop with the relevant properties (Quine 1956).

Irreducibly second-person thought is a species of singular thought. As such, an adequate descriptivist account of second-person thought must capture intuitions about its singularity. Whether or not descriptivism can adequately capture the singularity of thought more generally, in attempting to do so, descriptivism puts pressure upon the irreducibility of second-person thought. For example, suppose that the second-person mode of presentation amounts to something like "the agent with whom I am interacting." This simple descriptivist account doesn't work, as one might be engaged in interaction with multiple agents simultaneously, but have a second-person thought about only one of them. But while there are various ways to begin defending descriptivism in response to this kind of charge, most plausible seeming ways of fixing the description so as to yield a unique referent don't furnish an account of *irreducibly* second-person thought. If so, descriptivism effectively faces a dilemma in trying to capture the irreducibly second-person mode of presentation. If descriptivism fails to capture the singularity of irreducibly second-person thought, then it misses a key element of the second-person mode of presentation. But if descriptivism can succeed in capturing the second-person mode of presentation, it most likely does so by driving a reduction of the second-person mode of presentation to some combination of the first- and third-person modes of presentation. In so doing, it fails to capture the irreducibility of second-person thought. Though I suspect that many philosophers have rejected the irreducibility of second-person thought on something like these grounds, one might instead think that the problem is with descriptivism.[93] What one owes, then, is an account of the second-person mode of presentation as a non-descriptive mode of presentation.

---

[92] Thus, I am using the term Descriptivism to name a thesis about reference determination, and not merely about the "cognitive significance" attached to singular terms or concepts.

[93] I don't take myself to have argued against the descriptivist account of the second-person mode of presentation here, or for the claim that descriptivists can't furnish an account of irreducibly second-person thought. Rather, I've identified

I take non-descriptive modes of presentation to be mental files. The picture of mental files that I subscribe to goes as follows.[94] Subjects stand in various relations to objects. Some of these relations are epistemically-rewarding relations, which allow the subject to gain information about the object that she bears the relation to. Relations of acquaintance are paradigmatic epistemically-rewarding relations, and there are typically a number of ways in which a subject might be acquainted with one and the same object. Depending upon the cognitive and perceptual capacities of the subject, some epistemically-rewarding relations are such that there is a specific type of mental file that corresponds to them. These types of mental files are individuated by their function. In each case, the function of the mental file is tied to the epistemically-rewarding relation(s): files function to store information about an object acquired in virtue of an epistemically-rewarding relation(s). While this basic account needs some elaboration,[95] with it one can distinguish three aspects of a mental file: the file itself, the information content of the file, and the referent of the file.

Mental files are mental particulars, and, specifically, the vehicles of singular thoughts. File types are individuated functionally, as specified above, and serve to cluster information that is presumed to be about the same object. As such, one might expect to find distinct files for both distinct objects accessed through the same epistemically-rewarding relation, and for identical objects that are accessed through different epistemically-rewarding relations. This is desirable, as files perform the roles typically assigned to mode of presentation. For example, when I access one object through two different epistemically-rewarding relations, I form two files about the object. This allows the mental files framework to make sense of an individual holding certain patterns of belief without lapsing into outright irrationality. For example, if one has distinct files for Superman and Clark Kent, then one might rationally believe that Superman wears a cape, but deny that Clark Kent does, even if (unbeknownst to the thinker) each file refers to the same individual.

---

descriptivism as one plausible source of skepticism about second-person thought, and identified an alternative to rejecting irreducibly second-person thought: reject descriptivism.

[94] The account that follows agrees in large measure with the account of mental files offered by Recanati 1993, 2012, 2016. Recanati calls his account the "indexical model of mental files," and explains the account in terms of its similarity to the reference of indexical terms. I haven't followed suit here, thus I won't be defending any claims about the similarity between the account of mental files I develop here and the reference of linguistic indexicals. While Recanati's appeal to the semantics of indexicals is expositorily useful, it also has the potential to instigate disagreements that are unnecessary for my purposes here (e.g. Millikan 1990's claim that there are no "mental indexicals.").

[95] In particular, on Recanati's picture, it is not ultimately the case that each file is tied to a single epistemically-rewarding relation. Recanati distinguishes a few varieties of files, including encyclopedia entry files, which are based on a higher-order epistemically-rewarding relation—the relation of bearing one of epistemically-rewarding relations $R_1,\ldots R_n$ to an object.

So understood, mental files capture the "cognitive significance" of a thought about an object, and it is mental files that explain the distinctive pattern of inferences that thought licenses, and the pattern of actions (and interactions) that thoughts explain. Capturing this significance is to be understood in terms of the function of the file itself, and not the specific information stored in the file. Files store information in such a way that presumes it to be about a single object, and thus accomplish a kind of informational clustering. In the simplest case, a mental file stores all of the information about a single object O gained in virtue of an epistemically-rewarding relation R. More complex types of files store other information as well. While there are various ways of representing and accessing information in this kind of database, for the sake of simplicity, I'll follow Recanati in thinking of the information in mental files as a list of predicates that are presumed to apply to the referent of the file. They key point is that one can think about an object by using a file without thinking all of the things that are stored in the file. As such, the functional role of the file can remain stable in the face of changes to the information content of the file.

Finally, the referent of a mental file is not determined by satisfying the information content of the file, but by being related to the thinker via the epistemically-rewarding relation. In slogan form, the reference of mental files is determined "relationally, not satisfactionally" (Bach 1987, Recanati 2012). For example, suppose that I were perceptually acquainted with Bob, and on the basis of this acquaintance relation, I formed a mental file about him. As it turns out, Bob instantiates very few of the properties in the file, while someone else, Steve, instantiates all of them. The mental files still refers to Bob, because he is the agent with whom I am perceptually acquainted. So understood, mental files might mis-represent when the referent of a file doesn't instantiate the information content (e.g. the properties) contained in the file.

I'll be assuming this basic account of mental files in order to give an account of the second-person mode of presentation. While I've argued neither for the framework itself, nor the specific version that I'm using,[96] there is plenty to recommend it. It is particularly useful for current purposes, as it offers a means of addressing various philosophical puzzles about thought and language that lead philosophers to posit modes of presentation (e.g. Frege puzzles), but without adopting more complex

[96] Recanati's version of the mental files framework is particularly well-developed, but is only one of the available variants.

accounts of propositional content. That said, the chief virtue of the mental files framework is its utility in understanding the nature of thought. The mental files framework offers a means of addressing various philosophical puzzles about perception and cognition, and has been developed by a range of philosophers, including John Perry, Jerry Fodor and Francois Recanati, to give a broadly computational account of cognition. It has also been put to work in understanding perception, attention and memory (Kahneman and Treisman 1984; Triesman 1988; Kahneman, Triesman and Gibbs; Pylyshyn 2007; Carey 2009 and Greene & Quilty-Dunn forth), and various linguistic phenomena including word meaning (Gasparri 2016), anaphora (Karttunen 1976), and definiteness (Du Bois 1980; Heim 1983, 1988).

5.2 Two Varieties of Mental Files: Object Files and Self Files.

Before utilizing the framework myself, it is worth illustrating how types of mental files are distinguished by examining some commonly discussed examples. Doing so not only illustrates how the mental files framework is generally supposed to work, but offers a contrast between the types of files that others have previously been articulated (including object files and self files), and the new file type I describe in this paper (which I will call interlocutor files).

To begin, consider object files. Perceptual systems represent objects in the environment. An important element of doing so involves determining the persistence conditions of these objects (e.g. by determining when visual objects ought to be counted as stages of the same object, or as altogether different objects). One major account of how perceptual systems manage this task contends that they create object files to not only individuate and track particulars, but to store information gained about them. Thus, object files sit in between "low-level" sensory processing of color, shape, movement, etc. and "higher-level" forms of categorization. The object files account enjoys some empirical support,[97] and converges with philosophical work on the need for directly referential terms (Pylyshyn 2007).

Put into the mental files framework just introduced,[98] the object files account contends that files are a type of file, distinguished by their function. Object files serve to individuate and track particulars to

[97] For example, object files can explain certain visual effects like the object specific preview benefit (Kahneman et. al. 1992, Noles et. al. 2005), and provides a unifying framework for work on visual object perception and developmental work on the infant's "object concept" (Carey and Xu 2001, Scholl 2001, Carey 2009).

[98] It is worth noting that there is some variation between different mental files accounts, and that not all proponents of object files accept the version of the mental files framework that I introduced earlier.

whom one bears the epistemically-rewarding relation of perceptual acquaintance. Visual object files are files which individuate and track particulars to whom one bears the relation of *visual acquaintance*. This relation has a distinguishing feature: it follows the "principle of spatiotemporal priority" (Flombaum et. al. 2009). That is, the visual acquaintance relation is particular attuned to the spatiotemporal features of particulars, and not their qualitative features. This isn't to say that visual object files don't store qualitative information about particulars. Rather, it is to say that what we bear the relation of visual acquaintance to is informed primarily by spatiotemporal features of a given particular. In particular, the visual acquaintance relation is sensitive to continuity in the spatiotemporal features of an object (its location and movement), as opposed to continuity in the qualitative features of an object. Thus, if "objects" move in ways that violate certain types of spatiotemporal continuity, we tend to represent them not as a single object that "jumps around" but as different objects flashing in and out of existence. However, if objects exhibit spatiotemporally continuous movement, despite changes in qualitative features, we experience them as a single object that is changing features.

Despite their similarity, object files are not fully-blown mental files in the sense I introduced earlier. Mental files play the role of mode of presentation of the individual that they purport to be about, and are meant to be concepts (where to be a concept is, roughly, to be a thought constituent). Given that mental files play this role, they must meet certain constraints. For example, Murez and Recanati 2016 offers two such constraints:

> *Frege's Constraint*: "If m is a mode of presentation under which a minimally rational person x believe a thing y to be F, then it is not the case that x also believes y not to be F under m."
>
> *Trading-on-Identity constraint*: "From 'a is F' and 'b is G', one can rationally infer that some x if F and G only if 'a' and 'b' are associated with the same mode of presentation (not merely the same reference). If 'a' and 'b' are coreferential without being associated with the same mode of presentation, an additional premiss 'a=b' is required to support he inference to 'Ex Fx&Gx'."[99]

Object files don't meet these criteria, as they are limited in the properties that they can represent particular as having. Accordingly, object files are *proto-files*. As such, they represent an extension of the mental files framework that I introduced earlier. But they serve to illustrate some basic points about the functional role that files play in one's cognitive economy: they serve to cluster information obtained about a particular object, their referent is determined relationally, etc.

[99] Murez and Recanati 2016, p. 269-270. The Trading-on-Identity constraint was drawn from Campbell 1987.

Next, let's turn to self files. Self files are based upon the relation of identity that every individual bears to herself. Identity, it turns out, is an epistemically-rewarding relation: "in virtue of being a certain individual, I am in a position to gain information concerning that individual in all sorts of ways in which I can gain information about no one else" (Recanati 2012, p. 26). The self file functions to store information gained via the epistemically-rewarding relation of identity. Given that we only bear this relation to ourselves, we have only one self file.

The self file has two unusual features. First, the relation of identity is stable. Because the object that we are identical with doesn't vary by context, identity doesn't vary contextually, unlike many of the other epistemically-rewarding relations. This doesn't make the self file different in kind from other mental files, but rather it reflects a difference in the character of the epistemically-rewarding relation that this type of file is based upon. The lack of contextual variability of identity yields an unusually stable type of file. Second, the self-file illustrates a key component of genuine mental files, namely that they store two different kinds of information. The self file not only stores first-person information (gained in virtue of the epistemically-rewarding relation of identity), but also information "not gained in this way, but concerning the same individual as information gained in this way" (Recanati 2012, p. 66).

Further, the self file fits the other characteristics of the mental files framework introduced previously. Its reference is determined relationally, and not because the referent satisfies the content of the information in the mental file. Like other files, the self file is positioned well to play the role of mode of presentation. But unlike other files, the self file is a plausible candidate to play the role(s) of the irreducibly first-person mode of presentation operative in irreducibly first-person thought.[100]

These examples suggest a basic recipe for distinguishing types of mental file. With a distinctive function comes a distinctive file type. Thus, the first step to distinguishing a novel file type is to isolate the epistemically-rewarding relation of interest. Next, by appeal to this relation, articulate the function of the file type. Having functionally distinguished a type of file, one might then link it to a recognizable type of thought. Linking a file type to a recognizable type of thought requires meeting a set of intuitive

[100] This might be shown by, e.g., arguing that the self file has the explanatory and epistemic properties of irreducibly first-person thought (e.g. immunity to error, plays an ineliminable role in rationalizing action, etc.).

criteria. The referent of the thought type must be an intuitive match for the object that the agent bears the epistemically-rewarding relation to, as the cognitive significance of the thought type must match the function of the file type. We see this in each of the file types examined in this section. Further, there is no reason to think that the file types examined here are exhaustive. Thus, so long as the steps can be non-trivially duplicated, the recipe might be used again to distinguish novel types of files.

5.3 Second-Person Mental Files.

The main contention of this chapter is that we can distinguish a further type of mental file, which I'll call interlocutor files, with which we can plausible identify the second-person mode of presentation. The pattern of the argument is thus the same as the previous file types: upon isolating the relevant epistemically-rewarding relation, one can functionally define a new type of file. I'll also identify some core features of interlocutor files. Because the second-person mode of presentation shares these features, it is plausible that second-person modes of presentation are interlocutor files. Further, this account will *ipso facto* accommodate the irreducibility of second-person thought, as interlocutor files are based upon a different relation than the files affiliated with first- and third-person thoughts.

5.3.1 Direct Engagement as an Epistemically-Rewarding Relation.

The first step in identifying a distinct type of file is to specify the epistemically-rewarding relationship that its function will be specified in terms of. For current purposes, I'll call the relation in question *direct engagement*. To better appreciate what direct engagement is, and isn't, I'll contrast it with another epistemically rewarding relation, perceptual acquaintance.[101] When we are in good perceptual contact with an object, we bear the relation of *perceptual acquaintance* to it. In virtue of this relation, we might open a mental file about the object of our acquaintance that serves to cluster information about it. But there is a key epistemic difference between observing objects and interacting with other agents. Because mere objects don't perceive us back, or respond to our actions in any substantive sense, perceptual acquaintance is unrequited. This isn't a fact about our psychology, but about the channel of information about an object that perceptual acquaintance affords to a subject. Of course, the subject might respond differentially to objects depending on what they are doing, as when one is visually acquainted with a baseball flying towards their face. But one needn't (and indeed can't) interact with

[101] There are presumably species of perceptual acquaintance, e.g. visual acquaintance, auditory acquaintance, etc. What I say more generally about perceptual acquaintance in this section might be applied to each of these more specific acquaintance relations in turn.

mere objects, because they never change what they are doing in light of what you are doing. Thus, our reactions to objects are one sided—they involve us (as agents) reacting to the world around us.

When agents interact, things are not so simple. Agents are objects, but not mere objects. When we interact with agents, we might be perceptually acquainted with them, and we might intake information in about the agent (or more accurately, the agent's body) on the basis of the perceptual acquaintance. But, in the context of a face-to-face interaction, this isn't the only channel through which information about an agent might be conveyed to us. When one is engaged in a reciprocal interaction with another agent, in which your act is a response to their act, theirs a response to yours, etc., one might bear the epistemically-rewarding relation of *direct engagement* to the other agent with whom you interact. Such interactions tend to be temporally extended (to some degree), and afford us a further opportunity to intake information about other agents. Thus, in the context of a face-to-face interaction with another agent, I might bear the relation of perceptual acquittance to them qua-object or even qua-agent. But I might also bear the relation of direct engagement to them qua-*agent with whom I am interacting* (where interaction is of the reciprocal variety that I've been interested in).

One might worry that direct engagement is not a distinct epistemically-rewarding relation, and is instead simply a form of perceptual acquaintance. The worry here is not that direct engagement is a species of the broader genus of acquaintance relation.[102] Rather, the worry is that direct engagement and perceptual acquaintance are ultimately the very same relation, and thus cannot be used to legitimately distinguish interlocutor files from other types of mental files.

This worry can be dismissed by clarifying the relata and the formal properties of the two relations. The relata of the relations are different. Direct engagement holds between agents, whereas perceptual acquaintance can hold between two agents or an agent and an object. But they might also be distinct. In most actual cases, direct engagement entails perceptual acquaintance, and the two are quite hard to pull apart. But, the two might be distinguished by considering a conceivable, but non-actual, ability like telepathy. If telepathy isn't some kind of non-auditory hearing, then it looks like a form of "conversational" engagement without perceptual acquaintance. One might be engaged with an agent whom they are telepathically communicating with, but not be perceptually acquainted with them. If

[102] Following Lewis 1999, many philosophers use the term 'relations of acquaintance" as a genus of which there are more determinate species.

so, the conceivable cases of direct engagement are not merely a subset of the cases of perceptual acquaintance.

More tellingly, the two relations have different formal properties. Perceptual acquaintance is asymmetric: if I am perceptually acquainted with an object, it needn't be perceptually acquainted with me. But direct engagement is symmetric: if I am directly engaged in an interaction with another agent, then they are directly engaged in an interaction with me as well.[103] I conclude that perceptual acquaintance and direct engagement, while similar in certain respects, are ultimately distinct epistemically-rewarding relations.

Having distinguished the relations, one might wonder what evidence we have that agents ever instantiate relations of direct engagement with each other. Again, comparing direct engagement to perceptual acquaintance helps to illustrate the possibilities. There are (at least) three lines of evidence that lend support to the idea that we bear relations of perceptual acquaintance to objects: phenomenological evidence, transcendental considerations based in the possibility of knowledge, and related considerations of physiological plausibility. Broadly speaking, the same kinds of evidence can be used to support direct engagement.

First, direct engagement is phenomenologically different from passive observation. It feels different to engage with another agent than it does to merely observe them, and there is a distinctive manner in which we experience another agent when we take ourselves to be interacting with them. I'll call this kind of experience *recognition experience*. The experience itself is familiar enough. Suppose that, much to your chagrin, some well-abandoned philosophical view has recently become trendy again. You are in the hall with a colleague loudly despairing of this, when you realize that someone else nearby—a colleague who prominently defends the view!—is listening to you. They look up at you, and you realize that what was a two-person conversation has now been enlarged. You now experience the stranger as a party to the conversation.

[103] It is worth pointing out that in this section I am stating the conditions under which the epistemically-rewarding relation of direct engagement is instantiated between two agents. I have not yet stated the conditions under which one might entertain a second-person thought. This distinction matters, especially as it turns out that second-person thoughts presume that the relation of direct engagement holds between the thinker and the agent with whom they are interacting, but don't entail that it does. Section 4 takes up this point in more detail.

Recognition experience is different from experiencing someone as being spatio-temporally located in some way, or even experiencing them as being a potential conversational partner. Recognition experience represents someone as being a locus of agency who is engaged with me in interaction. And it provides defeasible evidence in favor of that person actually being engaged with me. Of course, it might turn out that they aren't: maybe they weren't paying attention, or maybe I am a brain-in-a-vat. But the situation is no different with respect to phenomenological evidence in favor of perceptual acquaintance. Both kinds of experiences offer a kind of defeasible evidence in support of an external world that we actually instantiate relations to, which our best scientific theories attempt to accommodate.

Second, there is some psychological evidence suggesting that we have the kinds of capacities required to pick up and make use of relations of direct engagement. These center on our ability to utilize information about the timing of our actions, and the actions of things around us to determine that they are interacting with us. To begin, consider our ability to detect agents. Suzy Johnson summarizes her work showing that infants as young as 12-15 months old treat things that either look like an agent (e.g. have a face) or behave in indicative ways as an agent—they treat them as if they can attend to things, have communicative abilities and have goals (Johnson 2003). One of the cues that infants use to perform agency detection is whether the things exhibits contingent interactions with other known agents. Contingency, as understood here, is a matter of timing—if a thing consistently responds to another agent's actions in the right temporal pattern, it contingently interacts. Humans, from infants to adults, are sensitive to such fine temporal cues, and use this information to determine whether or not to treat an object in their environment as an agent.

Of course, more than mere agent detection is needed for direct engagement—Johnson's work shows that infants can "detect agency" whether or not they are interacting with a thing, or merely observing it. But, we don't merely have the ability to detect contingent interaction. By combining some understanding of the timing of our own actions, with the ability to detect contingent interaction, what we have is the ability to detect when something is interacting contingently *with us*. Thus, not only can we detect contingent interaction, but, as I would put it, I can detect things that are interacting with me. The extent of our abilities to detect this, what our brains might do with such information, and the nature of the mechanisms and the computations by this takes place, are all apt for empirical

investigation. Neither is contingent interaction the only form of relevant empirical evidence.[104] But it would be unsurprising, in some sense, if social animals like us have not evolved capacities to make use of an information channel like direct engagement.

5.3.2 The Second-Person Mode of Presentation as Interlocutor File.

Having distinguished direct engagement from related kinds of epistemically-rewarding relation, we can now describe a type of mental file which is opened on the basis of this relation. *Interlocutor files* are a type of file which functions to store information about an agent taken in on the basis of direct engagement with that agent. These files presume (but don't explicitly represent) that this information is about a single agent to whom the thinker bears the relation of direct engagement, and function to cluster information about the agent. Once the epistemically-rewarding relation and function distinctive of interlocutor files are set, the remainder of the account plays out in the usual way. The reference of the file is not determined satisfactionally (e.g. the referent is the object which best satisfies the information in the file), but relationally (e.g. the referent is the object that bears the relation of direct engagement to the thinker of the thought). Thus, interlocutor files might misrepresent, as the information put into the file may or may not be accurate. But if an interlocutor file refers, it can only refer to an agent (as one cannot bear the relation of direct engagement to anything other than an agent).

Token interlocutor files are mental particulars that (potentially) refer to agents in the external world. They are typed functionally, and token interlocutor files store information about particular individuals with whom we interact. The information stored in these files need not be distinctive. It is typically gained in a specific manner (via the epistemically-rewarding relation of direct engagement), which presupposes a context: interacting with the agent. Finally, interlocutor files play the role of mode of presentation, and thus seek to capture the "cognitive significance" of the thought. The mode of presentation is captured by the functional role of interlocutor files together with the specific agent they are presumed to refer to. Thus, there are separate interlocutor files opened when one directly engage with different agents, or when one bears distinct epistemically-rewarding relations to the same agent.

[104] For example, work on gaze and joint attention might also brought to bear upon our capacity to detect direct engagement.

Thus, interlocutor files are a species of singular thought, but they have features which distinguish them from other more familiar types of mental files, and even accounts of second-person thought that have been discussed in the literature. Opening an interlocutor file that refers to an agent is not the same thing as merely representing something as an agent, as such files carry with them the presumption that the thinker is directly engaged with the object of the file. Nor is opening an interlocutor file the same thing as having a "you-thought" about an agent, where you-thoughts are thoughts whose apt expression involves using the second-person pronoun 'you'.[105] Interlocutor files are distinguished functionally, and not in terms of their apt expression. Thus, one might token an interlocutor file despite it being infelicitous to address someone. Thus, interlocutor files offer a further, novel way of understanding what is distinctive about second-person thoughts.

But how exactly are interlocutor files related to second-person thoughts? The core contention of this paper is that the second-person mode of presentation can be understood in terms of interlocutor files. Second-person thoughts represent their object via a second-person mode of presentation, and this second-person mode of presentation is to be understood in terms of opening an interlocutor file. Specifically I'll advance the following hypothesis: the second-person mode of presentation just is an interlocutor file, and thus an agent has a second person thought S about an agent A in virtue of opening up an interlocutor file about A.

I'll call this account the *Second-Person mode of presentation as Interlocutor File* account, or (SPIF). A full defense of (SPIF) would proceed by showing that it captures the key features of the second-person mode of presentation. For example, the second-person mode of presentation ought to capture the distinctive epistemic features of second-person thought, and that the second-person mode of presentation ought to capture the distinctive explanatory features of second-person thought. In previous chapters I've argued that only the irreducibly second-person mode of presentation is up to this job. Thus, in order to fully defend (SPIF), one could argue to the effect that it is the only (or best) means of accounting for the irreducibly second-person mode of presentation.

I'm not offering a full defense of (SPIF) here. Instead, I'm happy to leave it as a hypothesis ripe for investigation. It is an appealing account, but my primary interest is to use (SPIF) to show that general

[105] Chapter 2 examined such accounts.

skepticism about modes of presentation is no reason to doubt the irreducibly second-person modes of presentation. For there are accounts of modes of presentation on offer that don't make commitments that typically drive skepticism about modes of presentation, but that nonetheless accommodate and maybe even explain the irreducibility of the second-person mode of presentation. Thus, the basic argument retains its appeal regardless of how whether the mode of presentation itself ends up in the propositional content of the second-person thought, or in its vehicle, a la (SPIF).

To summarize, the account that I've sketched takes the following shape. Second-person thoughts deploy second-person modes of presentation to represent their object. I've offered an intuitive gloss on the second-person mode of presentation: it represents its object as an agent with whom I am engaged in interaction. Suppose that this captures the second-person mode of presentation as a non-descriptivist mode of presentation, which can be fruitfully understood as a mental file. Mental files distinguish the referent of the file, the file itself, and the information content of the file, and types of mental files are distinguished by their function (which is to cluster information about a given object gained in light of a specific epistemically-rewarding relation). Interlocutor files are a type of mental file about other agents to whom the subject of the thought bears the epistemically-rewarding relation of direct engagement. I then offered (SPIF) as a hypothesis about the relation between the second-person mode of presentation and interlocutor files.

5.4 Some Objections to the Mental Files Account.

While I haven't argued that the second-person mode of presentation must be understood in terms of interlocutor files, I've shown why some reasonable unease about modes of presentation ought not lead one to reject irreducibly second-person thought. Instead, one might endorse (SPIF), or some suitably similar (SPIF-y) account of the second-person mode of presentation. Doing so avoids the standard problems facing any account that places the mode of presentation in the content of the thought itself, and does so without (obviously) enacting a reduction of the second-person mode of presentation. But this is accomplished by employing a mental files account of non-descriptive modes of presentation, a framework which is not itself without detractors. So, how serious are the problems that face the mental files framework generally, and (SPIF) in particular?

The most stringent objections to mental files are general, and not specific to (SPIF). The main objections to the framework stem from its ambition: if the mental files framework is offering a general

account of singular thought, then it must account for all (and only) instances of singular thought.[106] Many philosophers contend that it doesn't. For example, some have denied that bearing an acquaintance relation to the referent of a thought is necessary for determining the reference of singular thought (e.g. Manley and Hawthorne 2012, Hall 2013).[107] Others have argued that some mental files employ something like a description to play governing roles (e.g. Goodman 2016). Finally, accounts of mental files face a family of problems concerning the identity of modes of presentation over time (or between people), which collectively get labelled the problem of "cognitive dynamics" (e.g. Evans 1981, Ninan 2014, Onofri 2014, Recanati 2016).

It is important to note that the claims being assessed in these general debates about mental files are stronger than anything I rely upon in this chapter (e.g., the biconditional claim: one has a singular thought about O iff one has a corresponding mental file about O). I've assumed only that a certain set of singular thoughts (second-person thoughts) can be treated within the mental files framework. Consequently, this isn't the venue for a full scale defense of the mental files approach. Accordingly, instead of responding directly these general objections, I'll focus on replying to forms of objection that bear specifically upon (SPIF).

There are two main clusters of objections to (SPIF). The first cluster centers on the appeal to the epistemically-rewarding relation of direct engagement. For example, one might worry that such a relationship isn't necessary for second-person thought, or that engagement lacks the stability required to distinguish a distinct type of file, or determine the reference of interlocutor files. I'll develop and reject these in turn.

Objection 1: Direct engagement with some agent A isn't necessary for having a second-person thought about A. Instead, merely believing that one is engaging with an agent seems to be sufficient for having second-person thoughts which purport to be about that agent. Examples might involve smoke, mirrors, and brains in vats, or simply result from one's attempts to interact with someone who isn't

[106] Rachel Goodman labels the claim that "one thinks a singular thought about an object O iff one employs a mental file to think about O" the *mental files conception of singular thought* (Goodman 2016, p. 437).
[107] Here the mental files framework takes sides in a larger debate about the role of acquaintance in singular thought. Philosophers who endorse some form of acquaintance condition include Donnellan 1979, Evans 1982, McDowell 1984, Bach 1987, Salmon 1987, Kaplan 1989, and Recanati 1993, 2010. Philosophers who deny that such a condition is necessary include Jeshion 2003, 2004, Sainsbury 2005 and Manley & Hawthorne 2012.

paying attention to them. In either case, the problem stems from the suspicion that believing that there is a single agent with whom one is directly engaged is sufficient for their being able to have second-person thoughts that are about that agent, whether or not one is actually so engaged.

In response, I'll point out that whether or not believing that one is directly engaged with an agent might be sufficient for having a second-person thought about that agent, this kind of belief can't be developed into a plausible necessary condition. For it also seems like can have second-person thoughts in the absence of such beliefs, or even in spite of them.[108] This results in a kind of intuitive standoff, which presents an important challenge to the mental files account that I've presented. But the real bite of the challenge rests on an unduly narrow understanding of the role that epistemically-rewarding relations play in specifying the function of mental files. I've defined interlocutor files functionally: they play the role of accumulating information about an agent which is acquired by direct engagement with that agent. Thus, interlocutor files cluster information by object (presumed identity of referent) and source (epistemically-rewarding relation). While it is natural to read this function as a necessary condition for being an interlocutor file, one needn't. As Recanati puts the point: "That the subject should stand in a suitable ER relation to some entity (the referent of the file) is a *normative requirement* corresponding to the function of the file. But there is no function without a possibility of malfunction. Since malfunctioning can always occur, there is no reason why a file could not be tokened even though the normative requirement is not met" (Recanati 2012, p. 63).

Thus, while it remains the case that the function of interlocutor files is to store information about an agent taken in on the basis of direct engagement with that agent, bearing the relation of direct engagement to the (presumed) referent of the file is not a necessary condition for a mental file's being an interlocutor file. For a mental particular can be a malfunctioning interlocutor file, without ceasing to be an interlocutor file. In this case, the file retains the function of an interlocutor file, and accumulates information that is presumed to be about a single entity, with which the subject is directly engaged. In the case of malfunctioning second-person files, there is no such entity, and the file fails to refer.

[108] Standard objections to descriptivism can be used to make this point. If we are appropriately causally connected to an object, we seem to be able to think about it even though we lack thoughts that uniquely pick out the object, lack thoughts about how we are related to the object, have mistaken thoughts about its nature or how we are related to it, etc.

Objection 2: Direct engagement with other agents is too short-lived and choppy to underpin the continuity of singular thought. For example, suppose that I am playing a "peek-a-boo"-like game with a baby. It seems like I can have second-person thoughts about the baby throughout the game, and surely these thoughts represent the baby under the same mode of presentation. (SPIF) doesn't deliver this verdict as epistemically-rewarding relations like direct engagement are too short-lived, resulting in our opening and closing mental files on the same object too quickly

In general, (SPIF) delivers the verdict that one can have different thoughts that co-refer throughout an extended interaction, but that there are changes in mode of presentation when there are changes, or breaks, in ones means of accessing an object. These changes in mode of presentation are not problematic, and are in fact desirable, so I see this as a feature of the account, and not a bug. The following is a plausible desiderata on an account of modes of presentation: posit separate modes of presentation when Frege cases can arise, but don't posit separate modes of presentation when Frege cases can't arise. As such, the mental files account seems to cut things just as finely as it ought to. The kind of scenario imagined is indeed "Frege-case-able," as I might learn something which I could state as "you are identical to you*." Because mental files play the role of modes of presentation, token mental files ought to be short lived when there are changes, or breaks, in ones means of accessing an object. It is what allows to mental files account to explain Frege cases in such scenarios.

This doesn't fully alleviate the worry that drives objection 2, but rather clarifies it. The worry isn't really about the direct engagement relation itself, but about generating and maintaining stable mental files. More precisely, it is about how to account for the dynamics of file transformations over time—how should we understand the construction of longer-lived types of mental files, and the dynamics of interlocutor files and these longer lived files.

This brings us into a second family of objections to the mental files framework that have to do with cognitive dynamics. These are important questions, but not ones that I intend to answer here. For while there is some tendency to assume that second-person thoughts should be long-lived, it isn't obvious why we should think that they are, and I don't see that anything is lost if they aren't. While self-conscious thoughts might exhibit a fair amount of stability, because of the stability of the epistemically-rewarding relation on which they are based, second-person thoughts need not exhibit comparable stability. If second-person thoughts play a role in explaining interaction, and our patterns

of interaction are choppy and short lived, then it seems like we might expect second-person thoughts to be less stable than either first- or third-person thought. So be it. An account is owed by the proponent of mental files here, but not an account of second-person thought itself.

In short, none of these objections ought to push one away from (SPIF). There is certainly work left to be done: the notion of function being deployed ought to be further clarified, a more satisfying account of the persistence conditions for interlocutor files is owed. But these strike me less as devastating objections to the account, and more as further avenues for developing it. Moreover, nothing that I've said is meant to suggests that proponents of second-person thought must endorse (SPIF). If one's intuitions strongly favor putting the ER-relation into the content itself, one might instead opt for an account on which the second-person mode of presentation features in the propositional content of the thought itself. This chapter has effectively assumed that placing the irreducibly second-person mode of presentation inside the content of a thought is the easy case for proponents of second-person thought, and shown that even when addressing harder cases, there is a defensible account of the second-person mode of presentation available. Given that (SPIF) also appears to be a productive account of the second-person mode of presentation—it fits in nicely with current research programs in both philosophy and the cognitive sciences, and suggests further lines of research—it offers an appealing account of second-person thought.

5.5 Conclusion.

Face-to-face interaction with another agent seems to differ from merely observing them. But what, if anything, does the relevant difference come to? I've argued that one key difference is epistemic. When observing another agent, we might be perceptually acquainted with them, or draw inferences about them. But in meaningful, reciprocal interactions, there is a distinctive manner in which humans with standard socio-cognitive capacities might come to have knowledge of other agents. When interacting with another agent, one can bear the epistemically-rewarding relation of *direct engagement* to the other agent(s) with whom they are interacting. This relation affords an additional source of knowledge about them which is available only in the context of such interactions.

Human socio-cognitive capacities utilize this distinctive source of knowledge to form thoughts about the agents with whom we have these meaningful, face-to-face interactions. In particular, in the context of an ongoing, reciprocal interaction, we might form second-person thoughts about the other agent.

A second-person thought represents its object as *an agent with whom I am interacting*, and this non-descriptive mode of presentation is best understood as a mental file. In reciprocal, engaged interactions, a subject can bear the epistemically-rewarding relation of direct engagement to the other agents with whom they are interacting, and, in virtue of standing in this relation to an agent, a thinker can open up a distinctive type of mental file about the other agent. So understood, the second-person mode of presentation is a distinctive type of mental file, an interlocutor file.

The account developed here certainly isn't the only possible account of irreducibly second-person thought. I've utilized a particular version of the mental files framework, which makes extensive use of the relation of *direct engagement*, which I've introduced and briefly motivated. While one might ultimately reject the account I've given, it has distinctive advantages. It readily accommodates second-person thought into a general computational account of cognition, and does so without accruing commitments to particularly burdensome propositional contents. Each of these advantages suggest that incorporating second-person thought into a more general account of social cognition won't require foundational reform in the cognitive sciences. Instead, second-person thought can be included into a suitably pluralistic account of mindreading, which might still be understood in (broadly) representational terms.

Despite its lack of revolutionary fervor, the account still assigns second-person thought distinctive, and interesting, features. Because interlocutor files utilize a distinctive source of information, they are poised to exhibit distinctive epistemic properties. Francois Recanati has put self files to work in understanding immunity to error through misidentification (e.g., Recanati 2012, p. 66). One might expect analogous moves to be available to understand when a second-person variant of immunity would hold. Further, because interlocutor files have a distinctive way of relating agents to one another in interaction, they are poised to play unique roles in explaining interaction. This opens up possibilities for explaining the psychological differences between agents who are merely observing one another, and those who are interacting. Thus, not only might there be affective differences between mere observation and meaningful interaction, but there might be cognitive differences as well.

Conclusion: Putting Social Cognition into Perspective

The previous chapters defend the irreducibility of second-person thought, and offer an account of it. A second-person thought is a thought that represents its object via a second-person mode of presentation. Irreducibly second-person modes of presentation are such that their cognitive significance cannot be captured by some combination of first- and third-person modes of presentation. Irreducibly second-person thoughts are thoughts that represent their object via an irreducibly second-person mode of presentation. The central contention of this dissertation is that the arguments widely accepted to establish the irreducibility of first-person thought can be extended to establish the irreducibly of second-person thought. Given the seeming plausibility of the arguments for the irreducibly of first-person thought, we are pulled towards accepting analogous arguments for irreducibly second-person thought as well.

I further argued that the second-person mode of presentation can be understand as a kind of mental file—an interlocutor file—which itself is to be understood in terms of the epistemically-rewarding relation of direct engagement. While this second-person as interlocutor file (SPIF) account is not the only way to understand irreducibly second-person thought, it is independently appealing, and should serve to placate philosophers who are uneasy about appealing to modes of presentation in crafting an account of thought.

I've made the case for irreducibly second-person thought with an eye towards addressing some main sources of skepticism along the way. The basic argument for irreducibly second-person thought doesn't hinge upon assumptions about the relation of experience to thought, or contentious claims about the relation of second-person language to thought, or even on adopting any particular account of the second-person thought. Instead, once proper adjustments are made, the case for irreducibly second-person thought is an extension of something that most philosophers already accept in some form: the case for irreducibly first-person thought.

So understood, the recipe for irreducibly second-person thought proceeds in (roughly) the following steps:

Step 1: Articulate a defensible account of the second-person mode of presentation. I argue that the second-person mode of presentations represents its object as *an agent with whom I am interacting.* This gloss treats the second-person mode of presentation in broadly functionalist terms, and steers clear of the tempting suggestion that second-person thought is to be understood in terms of paradigmatic instances of interaction like communicative interactions or cooperative interactions. Separating the recognition function of second-person thought from the address function of second-person language use avoids a major source of philosophical skepticism about second-person thought.

Step 2: Identify argumentative grounds for endorsing irreducibly first-person thought, extend these to support irreducibly second-person thought as well. There is no shortage of philosophical discussion of self-consciousness. Since at least the time of Aristotle, self-conscious thought was a major topic of discussion, and its importance was only magnified in Descartes, Kant and post-Kantian idealism. Amongst contemporary analytic philosophers, Hector-Neri Castaneda, John Perry and David Lewis (CPL) are credited with developing a widely influential tradition of arguments for irreducibly first-person thought. One can show that first-person thought cannot be reduced to third-person thought by showing that first-person thought, but not co-referential third-person thought, plays an irreplaceable role in rationalizing action, and has distinctive epistemic properties. I argue the by focusing upon the rationalization of interaction and some epistemic features of second-person thought, we can construct analogous arguments for irreducibly second-person thought. This establishes a conditional claim: if widely endorsed arguments in the CPL-tradition establish the irreducibility of first-person thought, then analogous arguments establish the irreducibility of second-person thought.

Step 3: Construct the argument(s) for second-person thought. Having established the just mentioned conditional, this step proceeds as a rather painless modus ponens. This *Basic Argument* for irreducibly second-person thought has one key advantage: in its simplicity, it assumes little about the nature of second-person thought. While philosophers agree widely that arguments in the CPL-tradition establish the irreducibility of first-person thought, they disagree amongst themselves about exactly what these arguments establish, and how these arguments work. The Basic Argument washes away much of this disagreement: so long as some version of the CPL-argument succeeds, an analogous version establishes the irreducibility of second-person thought.

Step 4: Identify defensible accounts of the second-person mode of presentation. This final step is defensive. The completed argument for irreducibly second-person thought might leave philosophers with cold feet. After all, much of the generality of the basic argument is gained by utilizing the notion of a mode of presentation in a permissive manner, and many philosophers are uneasy about this notion to begin with. What is owed, then, is a means of putting paid to the notion of mode of presentation. There are various ways of doing this, and I have pursued only one of them. I've argued that second-person modes of presentation can be understood as interlocutor files. One needn't accept this account in order to accept the basic argument for irreducibly second-person thought. But the account yields some distinctive advantages, including that it naturally accommodates irreducibly second-person thought into a computational account of cognition.

In addition to its broad appeal, this recipe for irreducibly second-person thought clarifies two main reasons why so many philosophers have either overlooked or rejected second-person thought. Things go wrong not at the end, but at the beginning. It seems quite natural to frame questions about second-person thought in terms of 'I'-thoughts and 'you'-thoughts. But this way of proceeding it not philosophically innocent, as the linguistic meanings of 'I' and 'you' are quite different in certain respects. While this way of understanding the second-person mode of presentation is typically offered as a part of the set-up of the issue, and not a conclusion that is argued for, it is nonetheless a theoretical choice that philosophers make when understanding second-person thought. Thus, if only tacitly, many philosophers have been endorsing what I called a communication-based account of the second-person mode of presentation. I've argued that philosophers have reason to make this choice differently, as there are more appealing (not to mention defensible) alternatives available.

This tendency to theorize about 'you'-thoughts is reinforced by the common philosophical practice of prising the subject apart from the rest of the world, and then finding that the first-person perspective (or subjectivity) poses unique philosophical problems. Approaching philosophical issues by way of this rift between the subject and the world (what I called the *Subject-Object Wedge*), makes extending the arguments for irreducibly first-person thought seem misguided. Why would we try to *extend* arguments which seek to establish the unique (and uniquely puzzling) features of irreducibly first-person thought? Further, the S-O Wedge makes the reduction of second-person thought seem quite natural. The basic idea is that once we have some form of indexical thought to "anchor" our perspective, all other thought can be understood as either third-person thought, or as relativized to this anchor point. Again, I've offered an alternative that is more conducive to second-person thought: if the fundamental features of interest are not attached to the first-person perspective specifically, but to personal perspectives more generally, then extending these arguments to second-person thought looks to be quite natural.

Throughout this dissertation, my focus has been on defending irreducibly second-person thought. But having both argued for it and gestured at an account, the next step is to put second-person thought to work. I've said less about the implications of irreducibly second-person thought, or the directions that future work on second-person thought might pursue. Developing these connections will ultimately bolster the arguments that I've given here, as the applications of second-person thought are at once both natural and philosophically surprising. The next section outlines three lines of research

that might be developed: from clarifying the role of second-person thought in accounts of social cognition, to articulating an epistemology of direct engagement, and developing connections to work on moral psychology, especially to accounts our moral responsibility practices and the second-person stance.[109] Having sketched these applications, I finish with some concluding remarks.

Moving Forward: What to do with an account of Second-Person Thought?

In chapter 1, I approached second-person thought via issues internal to accounts of social cognition. Specifically, I examined the *Interactionist Critique*, which contends that standard accounts of social cognition cannot account for the role that interaction plays in social cognition. I used these arguments as a means of seeing the work that an account of second-person thought might do, if it could be had. While second-person thought does indeed bear upon an adequate understanding of social cognition, its implications are more far reaching. Indeed, it is in large measure developing an account of second-person thought which further clarifies the widespread applications that human forms of social cognition might have for other philosophical issues. In this way, it is by putting the second-person perspective into social cognition that we put social cognition back into philosophical perspective.

(i) *Social Interaction and Pluralistic Accounts of Social Cognition*: Chapter 1 used the Interactionist Critique of Standard accounts of Social Cognition to begin a discussion of second-person thought. I suggested that the radical ambitions of many versions of the Interactionist Critique are misplaced, and that the available empirical evidence better supports a kind of pluralism about social cognition. Importantly, this pluralism indicates that social cognition is to include both the more familiar observational kinds of thoughts about other minds, and thoughts about the minds of agents with whom we take ourselves to be interacting.

Chapter 1 noted that an account of second-person thought, if it could be had, would provide an interesting response to this critical challenge. Having argued for, and developed, such an account, future work might put it to use in responding to the interactionist critique. The goal of such a reply is not merely to respond to the critique, but to better articulate the plurality of processes deployed in social cognition. In particular, second-person thought has the potential to show that fairly standard

[109] These certainly aren't the only lines of research which one might pursue. Equally interesting and important applications include articulating the developmental trajectory of second-person thought (e.g., which is prior first-person thought, or second-person thought?), or a phenomenological account of second-person thought (e.g., which articulates the experiential structures that one finds in meaningful interaction and their connection to thought, the self, and others).

accounts of social cognition might pair second-person thoughts with traditional stories about mindreading to distinguish mindreading during interaction from mindreading during mere observation. More generally, the project of integrating the account of second-person thought presented here into a computational account of social cognition is a potentially fruitful one.

(ii) *The Epistemology of Second-Person Thought*: While the issue of interaction is what first pushes us towards recognizing the irreducibility of second-person thought, an account of second-person thought does more than just respond to problems in existing accounts of social cognition. Indeed, I find that reflecting upon the distinctive epistemic features of irreducibly first-person thought clarifies a second source of arguments for irreducibly second-person thought. In my experience, philosophers find such arguments to be less intuitively compelling than the arguments I've given in chapters 2 and 3, though this seems to be driven largely by an implicit application of the S-O Wedge (e.g. contrasting self-knowledge with knowledge of the external world, and lumping second-person thought into the latter). Either way, working out the epistemic differences between irreducibly first- and second-person thought might further clarify the nature of each. And while there are surely epistemic differences between them, it remains to be seen whether these differences result from distinctive features of the first-person perspective itself, or from something else. To investigate this, we might pursue questions like the follow: Can we construct second-person variants of puzzling cases of self-locating thought (e.g. sleeping beauty)? Might second-person thoughts exhibit immunity to error through misidentification relative to the second-person pronoun?

My suspicion is that the same basic features of perspectival thought explain the epistemic features of both irreducibly first- and second-person thought. Any epistemic differences between the two would then result from differences in the level of contingency associated with the different epistemic sources recruited by first- and second-person thought. For example, suppose that a first-person thought formed on the basis of introspection enjoys of level of epistemic security that second-person thoughts formed on the basis of direct engagement don't. The reason why doesn't have anything to do with distinctive features of the first-person mode of presentation itself. Rather, it has to do with the nature of introspection. Because one can't introspectively access features of other minds, or non-existing minds, if a self-conscious thought is formed on the basis of introspection, it can't fail to be the case that the mind having the thought exists (or existed when the introspection took place). Thus, thoughts formed on the basis of introspection establish the existence of their object with a kind of necessity.

Many philosophers feel uneasy saying analogous things about second-person thoughts formed on the basis of direct engagement, though I must admit that I don't see relevant differences. Either way, the epistemology of direct engagement, especially as a source of knowledge of other minds, remains an interesting area for further work.

(iii) *Connections to Moral Psychology*: Finally, second-person thought might be applied to a range of issues in moral psychology. For example, consider our moral responsibility practices—our practices of holding agents to be morally responsible for certain of their actions. P.F. Strawson's influential account of these practices understands them largely in terms of interpersonal interactions (Strawson 1962). Strawson spoke of the participant attitude that we take up when we participate in the interpersonal interactions that give so much richness to our lives. This participant attitude is typically characterized in terms of our rich emotional engagement with others, often in the form of reactive attitudes like resentment, indignation, guilt, esteem, etc. While reactive attitudes are typically understood as emotions of some kind, second-person thought yields new tools for understanding the connection between the reactive attitudes and interpersonal interaction. Thus, instead of merely understanding reactive attitudes in terms of their features qua-emotions, we might also understand certain reactive attitudes as a type of irreducibly second-person thought.

Further, second-person thought is quite naturally taken to be related to the second-person stance. In chapter 1 I hypothesized that only thinking agents who can take others to be interacting with them, might take up the second-person stance. This hypothesis represents a link between second-person thought and broad issues in moral development and metaethics. Michael Tomasello has argued that the second-person stance plays a crucial role in moral development, and has tied the capacity for taking up this stance to various forms of collaborative activity (e.g., Tomasello 2016, 2019). Further, Stephen Darwall has argued that morality is to be understood as equal accountability, and that making or acknowledging claims upon one another presupposes a second-person authority (e.g., Darwall 2006). Findings of broad philosophical interest might result from mapping the implications of an account of second-person thought onto work on the second-person stance.

Cumulatively, these applications of second-person thought aim at developing a picture of the human agent, acting and thinking in a social world. While these applications might uncover an ineliminable kind of sociality in thought, epistemology and action, the account of second-person thought that I've

developed here offers a less radical way of understanding claims about human sociality. I find it natural to construe of agents with the capacity for second-person thought as being inescapably *embedded* in the social world in certain ways. But capturing these forms of embeddedness requires neither revisionary metaphysical claims (e.g. about the extended mind), nor sweeping methodological reform (e.g. widespread changes to the explanatory practices of the cognitive sciences). It does require recognizing certain limitations to attempts to study cognitive systems in isolation from their environment. But these aren't different in kind from limitations that practicing cognitive scientists and philosophers recognize already. Far from radicalizing accounts of social cognition, incorporating the second-person perspective into social cognition humanizes them.

Concluding Thoughts.

Robert Nozick begins his book *Philosophical Explanations* by admitting that he isn't engaging in what he calls "coercive philosophy." Though our beliefs might be demolished by philosophical argument, Nozick notes that this doesn't force us to change them: one might simply continue holding a belief irrationally, or simply disregard the arguments against it. Further, he wonders what function a "perfect" argument—one that leaves no choice but to endorse its conclusion—would really serve. Even at its most successful, philosophy doesn't seem able to force its conclusions upon the unwilling. Nor is this a nice way to treat someone.

I share Nozick's mistrust of coercive philosophy. This isn't a general mistrust of argumentation, nor is it to suggest that the arguments given don't ultimately matter. Instead, it is a matter of better appreciating what philosophical argument can do for us. At least until I change my mind, I think that the arguments of the previous chapters succeed, and I wouldn't have given them if I didn't think they did. But they don't force conclusions about second-person thought upon anyone. The basic argument for irreducibly second-person thought hinges upon one's prior acceptance of irreducibly first-person thought, and the desire to treat seemingly like cases alike. Either of these commitments can be jettisoned, and some philosophers suggest that we do so. Maybe thought simply doesn't manifest the kinds of personal perspectives that Castaneda, Perry and Lewis have suggested it does. It might instead be "our view on the world is not primarily a view *from a perspective*. Our beliefs and desires are not organized around us. They are instead organized around the world itself…Our view is a *view from everywhere*" (Cappelen and Dever 2013, p. 180). Alternatively, one's certainty that first- and second-

person thought are relevantly different could outweigh the reasons that I've offered for treating them alike. One needn't even reply, but can just "sit there silently, smiling, Buddhalike" (Nozick 1981, p. 4).

The basic argument might fail to compel belief among such interlocutors,[110] but its success is not contingent upon doing so in the first place. The argument succeeds because it regiments common philosophical commitments, and locates a choice point. Most philosophers haven't spared much thought to the irreducibly of second-person thought. Nonetheless, many of these same philosophers are committed to irreducibly first-person thought, often on the grounds offered by the Castaneda-Perry-Lewis tradition. As the same kinds of argument that compel belief in irreducibly first-person thought can be extended to second-person thought, what the basic argument offers philosophers is a choice: it doesn't push one to its conclusion, rather it tries to pull anyone who offers up a hand.

Irreducibly second-person thought itself offers something further. Philosophical work in metaphysics and epistemology is often oriented around a felt rift between the thinking, acting and experiencing subject, and the world that surround her. Second-person thought offers a way of loosening the grip of the dichotomy between subjective and object, and thus seeing around this *Subject-Object Wedge*, without merely collapsing the distinction. It suggests that the first-person perspective while real enough, is but one instance of perspectival thought. Second-person thought is another instance, and one that is much less prone to solipsism than the former. Thus, the appeal of second-person thought ultimately lies in what else it might allow us to see.

[110] As Nozick puts it, it might fail to offer "arguments so powerful that they set up reverberations in the brain" such that "if the person refuses to accept the conclusion, he *dies*" (Nozick 1981, p. 4).

References

Avramides, Anita (2001). Other Minds. Routledge.

Babb, Matt (2017). "The Essential Indexicality of Intentional Action," *The Philosophical Quarterly*, 66(264), pp. 439-457.

Bach, Kent (1987). Thought and Reference. Oxford: Clarendon Press.

Bach, Kent (2007). "Reflections on *Reference and Reflexivity*," in O'Rourke, Michael and Washington, Corey (eds.) Situating Semantics: Essays on the Philosophy of John Perry, pp. 395-424. MIT Press.

Bermudez, Jose Luis (2017). Understanding "I": Language and Thought. Oxford University Press.

Burge, Tyler (1977). "Belief de re," *Journal of Philosophy*, 74(6), pp. 338-362.

Campbell, John (1987). "Is Sense Transparent?" *Proceedings of the Aristotelian Society*, 88, pp. 273-292.

Cappelen, Herman and Dever, Josh (2013). The Inessential Indexical: On the Philosophical Insignificance of Perspective and the First Person. New York: Oxford University Press.

Carey, Susan (2009). The Origin of Objects. Oxford University Press.

Carey, Susan and Xu, Fei (2001). "Infant's' Knowledge of Objects; beyond Object Files and Object Tracking" *Cognition*, 80, pp. 179-213.

Carruthers, Peter (2011). The Opacity of Mind: An Integrative Theory of Self-Knowledge. Oxford University Press.

Cartwright, Nancy (2018). "Big Systems Versus Stocky Tangles: It Can Matter to the Details," *Erkenntnis*, 83, pp. 3-19.

Cassam, Quassim (2007). The Possibility of Knowledge. Oxford: Oxford University Press.

Castaneda, Hector-Neri (1966). "'He': A Study in the Logic of Self-Consciousness," *Ratio*, 8, pp. 130-157.

Castaneda, Hector-Neri (1967). "Indicators and Quasi-Indicators," *American Philosophical Quarterly*, 4, pp. 85-100.

Chisholm, Roderick (1981). The First Person: An Essay on Reference and Intentionality. University of Minnesota Press.

Cohen, Jonathan and Michaelson, Elliott (2013). "The Answering Machine Paradox," *Philosophy Compass*, 8(6), pp. 580-592.

Crane, Tim (2000). Elements of Mind. Oxford University Press.

Crossley, Nick (1996). Intersubjectivity: The Fabric of Social Becoming. SAGE Publications.

Darwall, Stephen (2006). The Second Person Standpoint: Morality, Respect, and Accountability. Cambridge, MA: Harvard University Press.

Devitt, Michael (2013). "The Myth of the Problematic De Se," in Feit, Neil and Capone, Alessandro (eds.) Attitudes De Se: Linguistics, Epistemology, Metaphysics, pp. 133-162. Stanford, CA: CSLI Publications.

Donnellan, Keith (1979). "The Contingent A Priori and Rigid Designators," Midwest Studies in Philosophy, 2, pp. 12-27.

Dretske, Fred (1973). "Perception and Other Minds," *Nous*, 7, pp. 34-44.

Du Bois, J. (1980). "Beyond Definiteness: The Trace of Identity in Discourse," in Chafe, W. (ed.) The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production, pp. 203-274. Norwood, NJ: Ablex.

Egan, Andy (2007). "Secondary Qualities and Self-Location," *Philosophy and Phenomenological Research*, 72(1), pp. 97-119.

Egan, Andy (2007). "Epistemic Modals, Relativism and Assertion," *Philosophical Studies*, 133, pp. 1-22.

Eilan, Naomi (2014). "The You Turn," *Philosophical Explorations*, 17(3), pp. 265-278.

Eilan, Naomi (2017). "You Me and the World," *Analysis*, 76(3), pp. 311-324.

Evans, Gareth (1981). "Understanding Demonstratives," in Parret, H. and Bouveresse, J. (eds.), Meaning and Understanding, pp. 280-304. Walter de Gruyter.

Evans, Gareth (1982). The Varieties of Reference. Oxford: Clarendon Press.

Feit, Neil (2008). Belief About the Self. New York: Oxford University Press.

Fodor, Jerry (1975). The Language of Thought. Harvard University Press.

Fodor, Jerry (1994). "Jerry Fodor," in Guttenplan, S. (ed.) A Companion to the Philosophy of Mind, pp. 292-299. Oxford: Blackwell.

Fichte, J.G. (2000). Foundations of Natural Right, according to the Principles of the Wissenschaftslehre. Cambridge: Cambridge University Press.

Flombaum, Jonathan; Scholl, Brian and Santos, Laurie (2009). " Spatiotemporal Priority as a fundamental principle of object persistence," in Hood, B.M. and Santos, Laurie (eds.) The Origins of Object Knowledge, pp. 135-164. Oxford: Oxford University Press.

Gallagher, Shaun (2001). 'The Practice of Mind: Theory, Simulation, or Interaction?" *Journal of Consciousness Studies*, 8(5-7), pp. 83-107.

Gallagher, Shaun (2004). "Understanding Interpersonal Problems in Autism: Interaction theory as an alternative to theory of mind," *Philosophy, Psychiatry, & Psychology*, 11(3), pp. 199-216.

Gallagher, Shaun (2009). "Two Problems of Intersubjectivity," *Journal of Consciousness Studies*, 16(6-8), pp. 289-308.

Gasparri, Luca (2016). "Mental Files and the Lexicon," *Review of Philosophy and Psychology*, 7, pp. 463-472.

Gillespie, Alex and Cornish, Flora (2009). "Intersubjectivity: Towards a Dialogical Analysis," *Journal for the Theory of Social Behavior*, 40(1), pp. 19-46.

Goldman, Alvin (2006). Simulating Minds: The Philosophy, Psychology and Neuroscience of Mindreading. Oxford University Press.

Gomes, Anil (2011). "Is There a Problem of Other Minds?" *Proceedings of the Aristotelian Society*, 111(3), pp. 353-373.

Gomes, Anil (2013). "Scepticism about Other Minds," in Machuca, D. and Reed, Baron (eds.) Skepticism: From Antiquity to Present, pp. 700-713. Oxford University Press.

Goodman, Rachel (2016). "Against the Mental Files Conception of Singular Thought," *Review of Philosophy and Psychology*, 7, pp. 437-461.

Green, E.J. and Quilty-Dunn, J. (forthcoming). "What is an Object File?" *British Journal for the Philosophy of Science*.

Haddock, Adrian (2014). "On Address," *Philosophical Topics*, 42(1), pp. 345-350.

Hall, J. Keith (2013). "Acquaintance and Mental Files," *Disputatio*, 5(36), pp. 119-132.

Heal, Jane (2014). "Second-Person Thought," *Philosophical Explorations*, 17(3), pp. 317-331.

Heck, Richard J. (2002). "Do Demonstratives Have Senses?" *Philosopher's Imprint*, 2(2), pp. 1-33.

Heim, Irene (1983/2002). "File Change Semantics and the Familiarity Theory of Definiteness," in Portner, Paul and Partee, Barbara (eds.) Formal Semantics: The Essential Readings, pp. 223-248. Oxford: Blackwell.

Heim, Irene (1988). The Semantics of Definite and Indefinite Noun Phrases. New York: Garland.

Horgan, Terry and Nichols, Shaun (2016). "The Zero Point and I," in Miguens, Sofia; Preyer, Gerhard and Bravo Morando, Clara (eds.) Pre-Reflective Consciousness: Sartre and Contemporary Philosophy of Mind, pp. 143-175. Routledge.

Hutto, Dan (2004). "The Limits of Spectatorial Folk Psychology," *Mind & Language*, 19(5), pp. 548-573.

Hutto, Daniel and Myin, Erik (2013). Radicalizing Enactivism: Basic Minds without Content. Cambridge, MA: MIT Press.

Jeshion, Robin (2002). "Acquaintanceless *de re* Belief," in Campbell, O'Rourke and Shier (eds.) Meaning and Truth: Investigation in Philosophical Semantics, pp. xx-xx. New York: Seven Bridges Press.

Jeshion, Robin (2004). "Descriptive Descriptive Names," in Reimer, Marga and Bezuidenhout, Anne (eds.) Descriptions and Beyond, pp. 591-612. Oxford: oxford University Press.

Johnson, Susan C. (2003). "Detecting Agents," *Phil. Trans. R. Soc. Lond. B*, 358, pp. 549-559.

Kahneman, Daniel and Treisman, Anne (1984). "Changing Views of Attention and Automaticity," in Parasumaran, R. et. al. (eds.) Varieties of Attention, pp. 29-62. New York: Automatic Press.

Kahneman, Daniel; Treisman, Anne and Gibbs, B. (1992). "The Reviewing of Object Files: Object-Specific Integration of Information," *Cognitive Psychology*, 24, pp. 175-219.

Kaplan, David (1989). "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals," in Almog, J., Perry, J. and Wettstein, H. (eds.) Themes from Kaplan, pp. 481-563. Oxford University Press.

Kaplan, David (1989). "Afterthoughts," in Almog, Joseph; Wettstein, Howard K. and Perry, John (eds.) Themes from Kaplan, pp. 565-614. New York; Oxford University Press.

Karttunen, L. (1976). "Discourse Referents," *Syntax and Semantics*, 7, pp. 363-385.

Kolbel, M. (2013). "The Conversational Role of Centred Content," *Inquiry*, 56, pp. 97-121.

Kukla, Rebecca and Lance, Mark (2009). Yo and Lo! The Pragmatic Topography of the Space of Reasons. Oxford University Press.

Lewis, David (1979). "Attitudes De Dicto and De Se," *The Philosophical Review*, 88(4), pp. 513-543.

Lewis, David (1983). "Individuation by Acquaintance and by Stipulation," *Philosophical Review*, 92(1), pp. 3-32.

Lewis, David (1999). Papers in Metaphysics and Epistemology. Cambridge: Cambridge University Press.

Longworth, Guy (2014). "You and Me," *Philosophical Explorations*, 17(3), pp. 289-303.

Mach, Ernst (1959). The Analysis of Sensations. Dover.

Magidor, Ofra (2015). "The Myth of the *De Se*," *Philosophical Perspectives*, 29, pp. 249-283.

Manley, David and Hawthorne, John (2012). The Reference Book. Oxford: Oxford University Press.

Marraffa, Massimo (2011). "Theory of Mind," *Internet Encyclopedia of Philosophy*. https://www.iep.utm.edu/theomind/#H3

McDowell, John (1983). "Criteria, Defeasibility, and Knowledge," *Proceedings of the British Academy*, 68, pp. 455-479.

McDowell, John (1984). "De Re Senses," *The Philosophical Quarterly*, 34, pp. 283-294.

McGinn, Colin (1983). The Subjective View. Oxford: Clarendon Press.

McKay, Thomas (1979). "On Proper Names in Belief Ascriptions," *Philosophical Studies*, 39, pp. 287-303.

Millikan, Ruth (1990). "The Myth of the Essential Indexical," *Nous*, 24, pp. 723-734.

Millikan, Ruth (2012). "Are There Mental Indexicals and Demonstratives?" *Philosophical Perspectives*, 26, pp. 217-234.

Millikan, Ruth (2018). Beyond Concepts. Oxford University Press.

Murez, Michael and Recanati, Francois (2016). "Mental Files: An Introduction," *Review of Philosophy and Psychology*, 7, pp. 265-281.

Nagel, Thomas (1986). The View from Nowhere. Oxford University Press.

Nichols, Shaun (unpub). "Mindreading and the Philosophy of Mind."

Nichols, Shaun and Stich, Stephen (2003). Mindreading. An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds. Oxford University Press.

Ninan, Dilip (2010). "De Se Attitudes: Ascription and Communication," *Philosophy Compass*, 5(7), pp. 551-567.

Ninan, Dilip (2013). "Self-Location and Other-Location," *Philosophy and Phenomenological Research*, 87(2), pp. 301-331.

Ninan, Dilip (2015). "On Recanati's *Mental Files*," *Inquiry*, 58(3), pp. 368-377.

Ninan, Dilip (2016). "What is the Problem of De Se Attitudes?" in Garcia-Carpintero, Manuel and Torre, Stephan (eds.) About Oneself, pp. 86-120. Oxford University Press.

Noles, N., Scholl, Brian and Mitroff, S. (2005). "The Persistence of Object File Representations," *Perception & Psychophysics*, 67, pp. 324-334.

Nozick, Robert (1981). Philosophical Explanations. Cambridge, MA: Harvard University Press.

Nunberg, Geoffrey (1993). "Indexicality and Deixis," *Linguistics and Philosophy*, 16, pp. 1-43.

Onofri, D. (2015). "Mental Files and Rational Inferences," *Inquiry*, 58(3), pp. 378-392.

Overgaard, Soren and Michael, John (2015). "The Interactive turn in Social Cognition Research: A Critique," *Philosophical Psychology*, 28(2), pp. 160-183.

Peacocke, Christopher (2014). The Mirror of the World. Oxford: Oxford University Press.

Perry, John (1977). "Frege on Demonstratives," *The Philosophical Review*, 86(4), pp. 474-497.

Perry, John (1979). "The Essential Indexical," *Nous*, 13(1), pp. 3-21.

Perry, John (1998). "Myself and I," in Stamm, Marcelo (ed.) Philosphie in Synthetisher Absicht, pp. 83-103. Stuttgart: Klett-Cotta.

Perry, John (2000). The Problem of the Essential Indexical and Other Essays, Expanded Edition. CSLI Publications.

Perry, John (2011). Reference and Reflexivity, 2nd Edition. CSLI Press.

Prosser, Simon (forth). "The Metaphysics of Mental Files," *Philosophy and Phenomenological Review*, xx, pp. xx-xx.

Prosser, Simon and Recanati, Francois (eds.) (2012). Immunity to Error Through Misidentification. Oxford University Press.

Pylyshyn, Zenon (2007). Things and Places: How the Mind Connects to the World. Cambridge: MIT Press.

Quine, W.V.O. (1956). "Quantifiers and Propositional Attitudes," *Journal of Philosophy*, 53, pp. 177-187.

Ratcliffe, Matthew (2006). "Folk Psychology is not Folk Psychology," *Phenomenology and the Cognitive Sciences*, 5(1), pp. 31-52.

Recanati, Francois (1993). Direct Reference: From Language to Thought. Oxford: Basil Blackwell.

Recanati, Francois (2012). Mental Files. Oxford University Press.

Recanati, Francois (2016). Mental Files in Flux. Oxford University Press.

Reddy, Vasudevi (2008). How Infants Know Minds. Cambridge, MA: Harvard University Press.

Reddy, Vasudevi (2018). "Why Engagement? A Second Person Take on Social Cognition," in Newen, Albert; De Bruin, Leon and Gallagher, Shaun (eds.) The Oxford Handbook of 4E Cognition, pp. 433-452. Oxford University Press.

Rodl, Sebastian (2007). Self-Consciousness. Harvard University Press.

Sainsbury, Mark (2005). Reference without Referents. Oxford: Oxford University Press.

Salje, Lea (2017). "Thinking about You," *Mind*, 125(503), pp. 817-840.

Salmon, Nathan (1986). Frege's Puzzle. Cambridge, MA: MIT Press.

Salmon, Nathan (1987). "How to Measure the Standard Metre," *Proceedings of the Aristotelian Society*, 88, pp. 193-217.

Schilbach, Leonhard (2010). "A Second Person Approach to Other Minds," *Nature Reviews Neuroscience*, 11(6), p. 449.

Schilbach, Leonhard; Timmermans, Bert; Reddy, Vasudevi; Costall, Alan; Bente, Gary; Schlicht, Tobias and Vogeley, Kai (2013). "Toward a Second-Person Neuroscience," *Behavioral and Brain Sciences*, 36(4), pp. 393-414.

Scholl, Brian (2001). "Objects and Attention: the State of the Art," *Cognition*, 80, pp. 1-46.

Schonherr, Julius (2017). "What so Special About Interaction in Social Cognition?" Review of Philosophy and Psychology, 8, pp. 181-198.

Schonherr, Julius and Westra, Evan (2019). "Beyond Interaction: How to Understand Social Effects on Social Cognition," *British Journal for the Philosophy of Science*, 70, pp. 27-52.

Shoemaker, Sydney (1968). "Self-Reference and Self-Awareness," *Journal of Philosophy*, 65, pp. 555-567.

Smith, Joel (2017). "Self-Consciousness," *Stanford Encyclopedia of Philosophy*.

Spaulding, Shannon (2015). "Phenomenology of Social Cognition," *Erkenntnis*,

Spaulding, Shannon (2018). How We Understand Others: Philosophy and Social Cognition. New York: Routledge.

Stalnaker, Robert (1981). "Indexical Belief," *Synthese*, 49(1), pp. 129-151.

Stalnaker, Robert (2016). "Modeling a Perspective on the World," in Garica-Carpintero, Manuel and Toree, Stephen (eds.) About Oneself: *De Se* Thought and Communication, pp. 121-137. New York: Oxford University Press.

Strawson, P.F. (1959). Individuals. London: Metheun.

Strawson, P.F. (1962). "Freedom and Resentment," *Proceedings of the British Academy*, 48, pp. 1-25.

Stroud, Barry (1989). "Understanding Human Knowledge in General," in Clay, Marjorie and Lehrer, Keith (eds.) Knowledge and Skepticism, pp. 31-50. Westview Press.

Taylor, Charles (1994). "The Politics of Recognition," in Gutmann, Amy (ed.) Multiculturalism: Examining the Politics of Recognition, pp. 25-73. Princeton University Press.

Teller, Paul (2011). "Review: Robots, Action and the "Essential Indexical"," *Philosophy and Phenomenological Research*, 82(3), pp. 763-771.

Thau, Michael (2002). Cognition and Consciousness. Oxford: Oxford University Press.

Tomasello, Michael (2016). A Natural History of Human Morality. Cambridge, MA: Harvard University Press.

Tomasello, Michael (2019). Becoming Human: A Theory of Ontogeny. Cambridge, MA: Harvard University Press.

Torre, Stephen (2010). "Centred Assertion," *Philosophical Studies*, 150, pp. 97-114.

Torre, Stephen (2018). "In Defense of *De Se* Content," *Philosophy and Phenomenological Research*, 97(1), pp. 172-189.

Treisman, Anne (1988). "Features and Objects," *Quarterly Journal of Experimental Psychology*, 40, pp. 201-237.

Weber, Clas (2013). "Centered Communication," *Philosophical Studies*, 166, pp. 205-223.

Wittgenstein, Ludwig (1953). Philosophical Investigations. Oxford: Blackwell.

Wittgenstein, Ludwig (1980). Remarks on the Philosophy of Psychology, Vol. 1 & 2. Oxford: Blackwell.

Zahavi, Dan (2007). "Expression and Empathy," in Hutto, D. and Ratcliffe, M. (eds.) Folk Psychology Re-Assessed, pp. 25-40. Dordrecht: Springer.

Zahavi, Dan and Overgaard, Soren (2013). "Intersubjectivity," in LaFollette, Hugh (ed.) The International Encyclopedia of Ethics, pp. 1-11. Blackwell.

Zawidzki, Tadeusz (2013). Mindshaping: A New Framework for Understanding Human Social Cognition. MIT Press.