October 2022

Naturalized Human Epistemology is Social Epistemology

Molly O'Rourke-Friel
University of Massachusetts Amherst

Naturalized Human Epistemology is Social Epistemology

A Dissertation Presented

by

MOLLY O'ROURKE-FRIEL

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2022

Philosophy

Naturalized Human Epistemology is Social Epistemology

A Dissertation Presented

By

MOLLY O'ROURKE-FRIEL

Approved as to style and content by:

_____
Dr. Hilary Kornblith, Chair

_____
Dr. Sophie Horowitz, Member

_____
Dr. Alejandro Pérez Carballo, Member

_____
Dr. Erik Cherise, External Member

_____
Phil Bricker, Department Chair Philosophy

DEDICATION

To my sister, Emma

ACKNOWLEDGMENTS

Though I am listed as the sole author, writing this dissertation has been a truly, madly, deeply social endeavor. I owe so much to so many, and what I write here will likely not do justice to the guidance, support, and feedback I have received.

I would like to thank the best advisor a graduate student could ask for: Hilary Kornblith. So much of who I am as philosopher and scholar is the result of his mentorship. The ideas in this manuscript were developed while I sat in the big red chair in his office, as he generously and enthusiastically discussed drafts of my work – drafts that were often sent to him just that morning or the night before. For your feedback on this dissertation, for investing in me as a philosopher, and caring about me as a person – thank you.

I also want to thank Sophie Horowitz and Alejandro Pérez Carballo for being wonderful committee members. From their co-taught Proseminar my first semester of graduate school, to our conversations about the chapters that follow, their feedback has been invaluable. I want to thank them for pushing me to be a better writer, for coaxing clarity and precision out of my work, and encouraging me to see the bigger philosophical picture. They did this while approaching our discussions with humor and joy; they made philosophy fun.

I am in debt to the entire UMass Amherst philosophy faculty, from whom I have learned so much in seminars, colloquia, and casual conversation. I want to say a special thank you to Ned Markosian, Vanessa de Harven, and Louise Antony.

The graduate students of UMass Amherst's philosophy department are owed an enormous thanks. Their philosophical engagement, friendship, and support were necessary not just for writing this dissertation, but making it through every step of the program. I want to recognize some folks who discussed my work at length, and whose friendship I value: Andréa Daventry, Sam Schechter, Victor Ma, Ryan Olsen, Daniel Haddad, Erin Weibe, David Turon, Tim Juvshik, Cruz Davis and Chaeyoung

Paek. Patrick Grafton-Cardwell and Marina Pérez del Valle are owed a special thanks for making South College E310 a happy and productive place to write this dissertation. This all said, the entire community has been integral. Thank you all.

Over the past couple of years, I have had the privilege of sharing many of the arguments included in this dissertation with the wider philosophical community. Their comments have been incredibly helpful. I want to thank the attendees of the *Why and How We Give and Ask for Reasons Conference* (University of Hradec Králové, 2021), my fellow panel members and the audience at the *Canadian Philosophical Annual Congress, Panel on Epistemic Normativity in Groups* (University of British Columbia, 2019), and participants of the *Cologne Summer School in Philosophy with Jennifer Lackey* (University of Cologne, 2018). I want to thank Katia Vavova, Josh DiPaolo, and Thomas Grundmann for reading and discussing my work. Thanks also to anonymous referees who have provided helpful feedback.

Finally, I want to thank the people outside of my UMass Amherst and philosophy bubbles, people who supported and cared for me while writing this dissertation.

Jeremy Levine's love saw me through to the finish line, and for that I am so grateful. For her constant friendship that kept me afloat, I thank Kate Shelton.

I want to express my deep gratitude to my family, for all they do. My parents' unwavering faith and confidence that I would finish this project buoyed me along. My sister discussed so many ideas with me, read so many paper drafts, that I am only somewhat joking – she probably has the equivalent of masters at this point. She did this all while providing care and reassurance. I am so thankful for all three of them: their love, encouragement, and patience.

ABSTRACT

NATURALIZED HUMAN EPISTEMOLOGY IS SOCIAL EPISTEMOLOGY

SEPTEMBER 2022

MOLLY O'ROURKE-FRIEL, B.A., MCGILL UNIVERSITY

M.A., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Hilary Kornblith

Our epistemic lives are ones of deep social dependence. Social epistemology is often understood as a subfield that stands apart from, but is compatible with, traditional individualistic approaches to epistemology. In my work I reject this view and argue instead that human epistemology is necessarily social epistemology. I argue for this as an epistemological naturalist. I understand epistemological naturalism as a commitment to the following: (a) the claim that empirical research from psychology, cognitive science, and evolutionary biology is relevant to epistemological inquiry and (b) the meta-epistemological thesis that knowledge and justification are reducible to natural phenomena. In Chapter 1 of my dissertation, I argue that a naturalistic epistemic lens can account for the phenomena and considerations that are foundational to non-naturalist arguments. This chapter not only defends epistemological naturalism from its opponent; it makes room for epistemic naturalists, reliabilists about justification in particular, to say that our social epistemic practices, like our ability to defend our beliefs to one another, are epistemically significant. Naturalistic reliabilists have historically just explained away non-naturalist intuitions about the importance of the human capacity for reasons-giving. This chapter gives naturalistic reliabilists the resources to claim that, while non-naturalists are mistaken when they characterize our reasons-giving capacity as the source of epistemic normativity, they are correct in thinking that it is of deep epistemic importance to creatures like us. In Chapter 2, I develop the naturalistic reliabilist theory of justification so that it can

accommodate empirical analysis of our social epistemic lives. I argue that there can be *extended interactive justification conferring processes*. In short, whether our individual beliefs are doxastically justified can be a function of the reliability of our dialogical interactions with interlocutors. My argument not only serves as a development and defense of reliabilism; it also functions as an independent argument against evidentialist views of doxastic justification. In Chapter 3, I turn my attention to epistemic blameworthiness and blaming. I give an argument against the plausibility of positing an epistemic norm of blameworthiness that is distinct from doxastic justification. I argue that internalists can't do so, because their notion of blameworthiness can't be meaningfully different from their notion of justification. I argue that it is difficult for externalists to do so, and that they ought not because of the fundamental commitments of externalism. I argue that we can give an account of our practices of epistemic blaming that construes them as instrumentally epistemically important. Chapter 1 establishes and defends a naturalistic social epistemic framework. The following two chapters explore how we should think about epistemic norms governing epistemic justification and blameworthiness if we adopt this framework.

TABLE OF CONTENTS

INTRODUCTION

I sometimes joke that social epistemology books and articles all start the same way or are at least pitched in the same register: "Since Descartes, traditional epistemology has been individualistic with respect to [*insert epistemic phenomenon here*]. I will argue that [*epistemic phenomenon*] is in fact deeply social..." Though hyperbolic, there is more than a grain of truth to my jest. I submit for your consideration the following:

> This book emerges out of my confidence in one core idea: the fact that we rely on others for so much of what we know about the world should prompt reconsideration of the individualistic orientation of traditional epistemology. (Goldberg 2010, 1)

> Until recently, epistemology—the study of knowledge and justified belief—was heavily individualistic in focus. The emphasis was on evaluating doxastic attitudes (beliefs and disbeliefs) of individuals in abstraction from their social environment. Social epistemology seeks to redress this imbalance why investigating the epistemic effects of social interactions and social systems. (Goldman and O'Connor 2021)

> Traditional epistemology, especially in the Cartesian tradition, was highly individualistic, focusing on mental operations of cognitive agents in isolation or abstraction from other persons. Roughly this traditional pursuit is what I have called *individual epistemology*. I have no general objection to individual epistemology…But given the deeply collaborative and interactive nature of knowledge seeking, especially in the modern world, individual epistemology needs a social counterpart: *social epistemology*. (Goldman 1999, 4)

> The social aspects of knowledge and justification are an important dimension of epistemology, and in the history of epistemology they have often been neglected. Recent work in social epistemology has partially filled this gap… (Audi 2006, 27)

> Despite the vital role that testimony occupies in our epistemic lives, traditional epistemological theories focused primarily on other sources, such as sense perception, memory, and reason, with relatively little attention devoted specifically to testimony. (Lackey 2006, 1)

> Social epistemology is a field of research within Anglo-American philosophy that has emerged over the last decades in the borderland between epistemology and philosophy of science. *Breaking with an ancient philosophical tradition*, social epistemology adopts a *social* perspective upon knowledge, construing it as a phenomenon of the public sphere rather than as an individual, or even private or "mental" possession. (Collin 2019, 21; emphasis mine)

I could go on, but I'll leave it there.[1]

This dissertation is on social epistemology. I was tempted to follow in the footsteps of my philosophical predecessors and write a similar introductory statement to scaffold the arguments I've crafted below. On reflection, I want to argue that this dissertation is making (or is at least trying to make) a greater break with tradition than this kind of characterization can do justice.

Social epistemology criticizes traditional epistemology for its myopic individualist focus. Usually, the criticism is that an oversight has been made: "There are epistemic phenomena that have been ignored, and we should make room for these phenomena in our epistemic theorizing". In short, we social epistemologists need our own sub-discipline. Consider Alvin Goldman's position in *Knowledge in a Social World*, quoted above. He writes that he "has no general objection to individualistic epistemology" (Goldman 1999, 4). The aim of his influential book is to make room for what has been overlooked. The contention is that this can be done without abandoning individualistic epistemic projects.[2]

Taken together, the chapters below are an argument for the claim that, with regards to human epistemology, social epistemology is not a sub-discipline; it is the whole ballgame.

My project starts by investigating the nature of epistemic normativity. I argue that we should take a naturalistic approach to epistemology and understand knowledge as a natural kind. This amounts to the view that knowledge is a natural phenomenon, something we can learn about and investigate through empirical research (not just armchair philosophy). Often,

[1] Though there is of a newer variant of this opener: "Social epistemologists have failed to appreciate how *thoroughly* social [*epistemic phenomenon*] is. In this paper, I will remedy this oversight…". For example, see Mark Alfano and Neil Levy's "Knowledge from Vice: Deeply Social Epistemology" where they write, "we believe that the account of epistemic agency that has emerged from the tentative turn to anti-individualism (Palermos 2016) in contemporary epistemology is inadequate" (2020, 888). See also Longino 2022.

[2] The Audi (2006) and Lackey (2006) quotes make this point clear as well.

committing to this view has meant rejecting the idea that there is a defensibility condition on knowledge, i.e., rejecting the idea that one must be able to say something in defense of one's beliefs to truly count as a knower. There are several reasons for this that will be discussed at length in Chapter 1. For now, I ask us just to consider that many epistemic naturalists argue that non-human animals and young children have knowledge. Given non-human animals and very young children are not capable of explaining why they take themselves to be justified in their beliefs, defensibility can't be a condition on knowledge.

While naturalists are usually comfortable rejecting the defensibility condition, it does come at a cost. Our ability to consciously reflect on our beliefs, to communicate why we think we should hold those beliefs, to question our interlocutor's credibility and evidence, all seem like important and epistemically significant parts of our human epistemic lives. Traditionally, naturalism has struggled to make room for this insight. In Chapter 1, I argue that we can account for the epistemic practices associated with defensibility through the naturalist lens. This involves seeing defensibility practices as a species-specific reliability-conducive epistemic practice. The privileged human environment, the environment humans must epistemically navigate, is sociocultural. Knowledge writ large isn't a function of our human social practices. But gaining knowledge as a human requires using cognitive faculties adapted for, and that are deployed in, social contexts. To oversimplify: when we are theorizing about human epistemology, we are engaged in social epistemology because humans are social creatures. Human epistemology is social creature epistemology, therefore human epistemology is social epistemology.

Chapter 1 vindicates a naturalistic, social methodological approach to human epistemology. Chapter 2 uses that methodology to investigate epistemic justification. I argue that analysis of empirical social psychological research suggests that, in cases where our beliefs

are formed through dialogical deliberation, the process that confers justification on our beliefs can extend beyond our individual cognition. In such cases, the relevant justification-conferring process includes the interactive exchange with our interlocutors. What is important about this thesis is that it highlights a kind of social epistemic dependence that has not previously been given sufficient consideration: an *interactive* kind of social epistemic dependence. There are certain human epistemic capacities that require dialogical social engagement to function. This means we must turn away from, or at least expand on, the following attitude towards social epistemology: Human epistemic subjects have their own individual knowledge-producing capacities that output beliefs. There is also a significant social aspect of human epistemic life because we get to gather the knowledge outputted by our community members' knowledge-producing capacities. Chapter 2 shows what is problematic about this attitude. We have knowledge-producing capacities that require dialogical social interaction. The social part of our epistemic lives is not just a kind of "bonus feature". Rather, it is part-and-parcel of the kind of epistemic creatures we are.

Finally, I turn to the notion of epistemic blame. A robust commitment to epistemic blame involves the following powerful intuition: "[W]hen we blame someone for his actions we are not merely saying it is bad that they happened, or bad that he exists: we are judging *him*, saying he is bad, which is different from him being a bad thing" (Nagel 1976, 322). At the core of this intuition is the assumption that there will be something philosophically interesting and normatively significant left when we isolate an individual from their environment and external forces. Though Chapter 3 has a broad scope, here I want to highlight my argument that undermines this assumption. Epistemic normativity is a function of the fit between an epistemic subject and her environment. As such there isn't anything epistemically significant or interesting about considering her in isolation. This means that we must do away with, or at

least seriously revise, our notion of epistemic blameworthiness. Some might be uncomfortable with this result, but we ought not be. Our epistemic lives are not just a function of our individual cognitive epistemic dispositions. Our epistemic lives are also about our environments, and the human environment is distinctively sociocultural. Chapter 3 is a demonstration of what it looks like to accept the conclusion that human epistemology is social epistemology – that it is not merely a sub-discipline that can leave individualistic epistemic theorizing relatively untouched.

CHAPTER 1

HUMAN KNOWLEDGE: ALL NATURAL, CERTIFIED ORGANIC

1. Introduction

There is a set of epistemological views bound by a shared commitment to the following thesis: while (human) knowledge has a necessary reliability condition, robust (human) knowledge also has a defensibility condition that is irreducibly epistemic and normative. By irreducibly normative, I mean to say that proponents of such views understand the defensibility requirement as inexplicable in non-epistemic, non-normative terms (such as reliability). Some understand the defensibility condition in terms of public epistemic norms: one has knowledge only if one can defend one's reliably formed belief by providing reasons in support of that belief. Michael Williams (2004, 2008, 2015), Edward Craig (1990), and Robert Brandom (2000) subscribe to such views, and contend that epistemic norms require that human knowers participate in social epistemic practices. Other epistemological positions in this set understand the defensibility condition as entailing that certain private, reflective intellectual capacities be exercised in the formation or evaluation of one's reliably formed belief. Such views posit that epistemic normativity is grounded in the exercise of certain intellectual virtues that satisfy a defensibility condition insofar as doing so demonstrates an accomplishment on the part of an epistemic subject.[3] Proponents of a private normative defensibility condition include Ernest Sosa (1991, 2007, 2015), Linda Zagzebski (1996, 2012), and Laurence BonJour (1985). Though the particularities of their views vary, the epistemologists in this camp understand epistemic normativity as a unique, irreducible kind of

[3] Sosa eloquently describes this position as "the view that knowledge is true belief out of intellectual virtue, belief that turns out right by reason of the virtue and not just by coincidence" (1991, 277).

normativity that cannot be broken down into non-normative, non-epistemic terms. Let us call the general position unifying this set of views *epistemic normativism*.

Such views stand in opposition to purely reliabilist accounts of human knowledge. Strict process or causal reliabilists, such as Alvin Goldman (1967, 1979) and Hilary Kornblith (2002), understand epistemic normativity as reducible to non-normative, non-epistemic evaluative standards. Indeed, in formulating an account of epistemic justification, Goldman explicitly states the following desideratum: "I want a theory of justified belief to specify in non-epistemic terms when a belief is justified" (1979, 1). Epistemic normativity, the pure reliabilist argues, is grounded in causal facts regarding the process by which an epistemic subject acquires her belief. As such, a theory of knowledge can ultimately be "couched in non-epistemic language" (Goldman 1979, 2). While an epistemic subject's ability to say something in defense of her beliefs may be an important way in which members of her epistemic community can come to identify what she knows, an epistemic subject's ability to defend her belief (either publicly or privately) does not bear on what she in fact knows. To emphasize the way in which the pure reliabilist takes epistemic normativity to be reducible to evaluative standards of reliability, I will (following Williams) refer to this epistemological position as "*austere reliabilism*" (Williams 2015, 251).

Epistemic normativism and austere reliabilism tend to be part and parcel of certain packages of epistemological commitments. The latter is typically associated with unyielding externalism, while proponents of the former often take themselves to walk a middle ground that incorporates the best of both internalist and externalist insights. The latter is consistent, arguably continuous, with epistemic naturalism and the claim that knowledge is a natural kind. The former typically goes hand in hand with the position that the irreducible normativity of

human knowledge precludes the category of knowledge from the status of natural kind.[4] This second distinction represents a particularly substantive divide between the two epistemological camps. An epistemologist's decision as to whether she takes the category of knowledge to be an institutional or natural kind informs her view on what constitutes the central epistemological project. Epistemic naturalists and non-naturalists disagree about the proper methods we can use to learn about knowledge; the role of metacognition in human epistemic practice; and how (if at all) human knowledge is meaningfully different from, or more "special" than, non-human animal knowledge.[5]

Kornblith (2002) is the most ardent defender of the claim that knowledge is a natural kind. Williams (2004, 2015) offers an argument on behalf of epistemic normativists against this claim. Williams' challenge to epistemic naturalism is available to any epistemological position that places a defensibility condition, public or private, on human knowledge. As such, it poses a serious threat to epistemological naturalism. While Kornblith (2011) has offered a reply to Williams' challenge, I expect epistemic normativists remain unsatisfied. It is the project of this paper to offer a more robust reply on behalf of the epistemic naturalists broadly, and the austere reliabilist more specifically, all the while welcoming the epistemic normativist's insight that defensibility is epistemically valuable. The current debate has established a false dichotomy between the following two options:

(a) Accept the defensibility condition on human knowledge. Argue that epistemic normativity is unique and irreducible to natural, non-epistemic standards. Find ways

[4] To clarify: to claim that human knowledge is not a natural kind or phenomenon is not to claim that knowledge doesn't occur in the natural world. I don't mean to be attributing any kind of mysticism about knowledge. Rather, I just mean to describe a position analogous to the (reasonable) view that the game of chess is an "institutional" as opposed to natural phenomenon despite being played in the natural world (Williams 2004, 194). More clarification of what it means to say that something is (or is not) a natural kind or phenomenon is provided below (Section 2).

[5] By "metacognition", I mean the capacity to think about one's thinking. Throughout this paper, "human metacognitive capacities" will be used interchangeably with "human capacities for reflective deliberation" (consciously entertaining and evaluating one's beliefs), and "reliability knowledge" (knowledge about the reliability of one's belief).

to argue for epistemic norms that are meaningfully distinct from standards of reliability.

(b) Accept that epistemic normativity is reducible to natural non-epistemic standards. Argue that a defensibility condition independent of reliability standards is implausible. Give no significant or privileged role to the (public or private) capacity for reasons-giving in developing a concept of human knowledge, or of the status of human knower.

Epistemic normativists are correct in claiming that defensibility, or the capacity to provide reasons for belief, is epistemically valuable. However, I argue they have misunderstood this epistemic value insofar as they locate defensibility as a necessary condition on knowledge. I pose an objection to epistemic normativists that draws from recent evolutionary and social psychological research to argue that the (public and private) normatively-laden practices they take to fulfill the standards of a defensibility condition can be understood in broadly naturalistic terms. Moreover, the results of this empirical research suggest that these normatively-laden practices are ultimately explicable in the non-normative, non-epistemic language of reliability. Defensibility, I will argue, is constitutive of species-specific reliability-conducive epistemic practices adapted for the privileged human environment. Once defensibility is given its proper place and understood through the naturalist's lens, we see that epistemological theories that account for defensibility are consistent with, and indeed support the claim, that knowledge is a natural kind.

2. Natural kinds

The full force of the debate under discussion requires a preliminary sketch of what it means to say that knowledge is, or is not, a natural kind. Natural kinds are mind-independent insofar as their existence is not contingent on the way in which humans understand them, the way in which they are used in human practice, or human specification of individuating

features.[6] The individuating features of a natural kind cluster together in the natural world independent of our particular interests and interventions. For example, gold would be exactly as it is even if humans did not know what gold was, or how it is distinguished from other kinds of elements.[7] Its kind-defining features – being a metal composed of atoms with seventy-nine protons – would be the same even if humans did not know about, or have a use for, gold, atoms, and protons. Gold would still be different *in kind* from pyrite (fool's gold) even if humans could not distinguish between the two and used both for the same purpose. What individuates gold from pyrite is not our understanding of these metals, the purposes to which they are put in human practice, or our positing of gold and pyrite kinds individuated by particular sub-atomic features. Rather, they are what they are because of the way the world is distinct from the "projection of" human interests (Kornblith 2011, 48).

To clarify, a kind is not natural in virtue of its individuating features being themselves natural phenomena or mind-independent. Humans could posit an institutional, quasi-elemental kind based on the presence of certain arbitrary subatomic features. For example, two high schoolers assigned a project on the periodic table of elements could divide up the work so that one student was responsible for researching elements with an odd number of protons, the other those elements with an even number of protons. In doing, they posit two categories: "odd-elements" and "even-elements". The central individuating subatomic feature of "odd-elements" and "even -elements," an atom's number of protons, is itself a natural, given phenomenon. However, the categories "odd-element" and "even-element" are not natural insofar as the member elements of those sets are not nomologically clustered together

[6] Both Williams and Kornblith would be amenable to this characterization of natural kinds. As Kornblith explains, what makes a natural kind "the kind of thing that it is...is a feature of the world itself, not a product of human ways of thinking" (2011, 44). Williams similarly claims that natural kinds "are independent of us: given phenomena" (2015, 252).

[7] I borrow this example from Saul Kripke's *Naming and Necessity* (1980).

in the natural world in a manner that is independent of human interest. All the elements with an even or odd number of protons do not share chemical properties in the way that Group 1 alkali metals do. As such, the categories "even-elements" and "odd-elements" are not natural kinds, but instead social or institutional.[8]

The ontological status of natural kinds, like gold, is distinct from the ontological status of anthropogenic or institutional kinds, like games. Games are anthropogenically mind-dependent: a game exists because *we constructed it*. Chess is the particular game it is because of the way in which we understand chess; the way in which we play the game. Absent human formulation and understanding of chess and its rules, nothing in the world would have the property of being in check. The property of a particular chess piece, the king, being in check depends on our specification of rules constraining the movement of the chess pieces, and the stipulated objective of the game. Importantly, these rules and objectives are, while non-arbitrary, relative to human interests. The rules of chess are not to be "discovered" in the natural world as they do not exist independent of particular human interests.[9]

In sum, anthropogenic or institutional kinds are mind-dependent in a robust sense. The individuating features of an anthropogenic or institutional kind are clustered together as the result of human stipulation. The cluster of individuating features of anthropogenic kinds is dependent on particular human minds, human understanding, for its existence. In contrast, natural kinds are mind-independent in a robust sense. A natural kind exists as an individuated kind independent of human comprehension of the individuating features of that kind.[10]

---

[8] Many thanks to Maya Eddon and Alejandro Perez Carballo for prompting me to clarify this point.

[9] As Williams makes clear, although social kinds "are what we say they are", they are not arbitrary insofar as they "are adapted to natural conditions and social ends" (2015, 253). The game of chess and its rules may be stipulative – or just what we say they are – but are not arbitrary insofar as they were crafted with the non-arbitrary aims of creating an interesting and enjoyable means of developing tactical and strategic skills.

[10] For the remainder of the paper, mind-dependent and mind-independent will refer to those concepts in the robust sense described here.

The central question can now be reframed as the following: is the category of human knowledge mind-independent, or mind-dependent? Is having knowledge more like the property of being gold or being in check?[11] The epistemic normativist wants to say that knowledge has an irreducibly normative dimension that makes it akin to the latter; the epistemic naturalist sees it as entirely like the former.

3. The Epistemic Normativist's Argument

I will now turn to outlining the epistemic normativist's argument, most clearly articulated by Williams (2004, 2015), for the conclusion that the category of (human) knowledge is not a natural kind. The argument rests on commitments regarding the normativity of knowledge. The epistemic normativist's argument, proceeding in two parts, is as follows:

> **(P1)** Evaluative standards of reliability do not constitute a normative condition on knowledge.
> **(P2)** There is a normative condition on knowledge because knowledge claims are not purely descriptive.
>
> > **(C1)** Knowledge has a normative condition independent of any reliability condition.
>
> **(P3)** If knowledge has a normative condition, then knowledge is not a natural kind.
>
> > **(C2)** Knowledge is not a natural kind.

After more fully developing this argument (in the remainder of Section 3), I will argue that current naturalist rebuttals that target (P3) (Kornblith, 2011) are likely unsatisfying to the epistemic normativist.

[11] Both Williams and Kornblith frame the debate in this way. In Williams 2015, we are asked if the concept of knowledge is "more like *water* or more like *offside*?" (251). Kornblith (2011) approaches the question by evaluating the way in which knowledge is similar to, and different from, water – a clear example of a natural kind.

3.1 (P1) - Standards of reliability are not normative

As was mentioned previously, many epistemic normativists take an advantage of their theory of knowledge to be the way in which it walks a middle ground between internalist and externalist theories of justification. As such, many epistemic normativists grant what they take to be the main insight of the externalist position in claiming that there is an externalist reliability condition on knowledge. Williams advocates for an account of "mature human knowledge," the permissible ascription of which is understood along two "dimensions": reliability and epistemic responsibility (2015, 249, 257). The former evaluative dimension, reliability, Williams himself describes as showing a "willingness [on his part] to recognize an element of truth in externalist reliabilism" (1999, 186). Sosa similarly has a bi-level theory of human knowledge that distinguishes between "animal knowledge" that a subject holds "with little or no benefit of reflection or understanding", and "reflective knowledge" that requires a subject not only to know what she knows, but know how the belief came about (1991, 240). Animal knowledge, understood as merely "apt" belief, is a precondition of reflective knowledge (Sosa, 2007, 24). Insofar as "aptness" is understood as the accuracy due to adroitness, it can be understood as a basic reliability condition (2007, 21). A reliable process produces beliefs that are likely to be true (accurate) because of its "epistemic competence" (adroitness) (2007, 21). However, neither Sosa nor Williams takes their reliability condition to fully establish the most developed – and most interesting – kind of human knowledge. Why is this the case?

Williams makes a familiar distinction between conforming to a rule and "acting in light of the rule" (2008, 11). Norms, by Williams' lights, necessarily play an explanatory role in a subject's intention and motivation. Therefore, they are not standards to which we merely conform, but rather rules that require our understanding and respect. A reliability condition

on knowledge places no such motivational or explanatory obligations on an epistemic subject's actions. One could conform to a standard of reliability without "acting in light" of a reliability rule. An epistemic subject can have "animal knowledge" without being motivated by a desire to act in accordance with standards of reliability. An epistemic subject's possession of "animal knowledge" can be explained purely with reference to the operation of particular cognitive processes in environments where their operation is reliable. In contrast, if one is to be capable of defending one's belief (publicly or privately, at least in some contexts), one needs to be motivated by certain norms: defensibility norms. Acting "in light of" defensibility norms allows such norms to explain one's epistemic successes.

3.2 (P2) - Knowledge is necessarily normative because knowledge claims are not purely descriptive.

While the above discussion establishes the epistemic normativist's claims that evaluative standards of reliability do not in fact amount to a normative condition on knowledge, it does not offer independent reasons for thinking that a theory of knowledge is lacking in the absence of any such normative condition. There are two main sources of evidence supporting the claim that epistemological theories need posit a normative condition on knowledge. The first concerns the question of epistemological interest. The epistemic normativists argue that we care about knowledge in a way that is distinct from the way in which we care about true belief.[12] We care about the reliability of a person's belief, yes, but also the "trustworthiness" of the agent's judgement (Sosa 1991, 240). To know that an epistemic subject is trustworthy, the epistemic normativist argues, an agent must (at least in some contexts, publicly or privately)

[12] Before introducing the distinction between animal and reflective knowledge, Sosa is careful to emphasize the following: "…we take an interest not only in the truth of beliefs but also in their justification" (1991, 240). Williams makes a similar claim:

> Epistemic norms are explained by reliability. Obviously, we have an interest in having true beliefs, for having true beliefs is important for achieving our purposes, whatever they happen to be. But why do we have an interest in knowledge, if knowledge involves more than true belief? At least part of the answer to this question is that beliefs are typically acquired. (Williams 2004, 207)

be capable of defending her belief. An epistemic agent is trustworthy insofar as she constrains her epistemic behavior in accordance with epistemic norms.

The second source of evidence for the claim that knowledge necessarily has a normative condition is the way in which ascriptions of Knowledge are modally governed. Knowledge claims, the epistemic normativist argues, are not just descriptive statements about the way in which the world, objectively and in fact, is. Rather, knowledge claims express responsible epistemic conduct, i.e. the extent to which a person's epistemic behavior fulfills "epistemic obligations" (Williams 2015, 253).[13]

If externalist standards of reliability do not place a normative condition on knowledge, and knowledge claims are expressive of the perceived fulfillment of normative obligations, then, the epistemic normativist concludes, (C1) is true: there is a normative condition on knowledge independent of any externalist reliability condition. I turn now to the epistemic normativist's claim that placing a normative condition on knowledge precludes the possibility that knowledge is a natural kind. This final stage of the argument rests on two claims. The first is that the modal governance of knowledge claims described above is incompatible with the thesis that the category of knowledge is a natural kind. The second is that the normative condition renders the category of human knowledge robustly mind-dependent in the manner of social or institutional, as opposed to natural, kinds.

---

[13] Note that the epistemic normativist's claim is that knowledge ascriptions are taken to be both descriptive *and* expressive.

Bonjour articulates his central notion of justification in responsibilist terms. "To accept a belief" without having good reason to think the belief is true, he claims is *epistemically irresponsible*": "My contention here is the idea of avoiding such irresponsibility, of being epistemically responsible in ones believing, is the core notion of epistemic justification" (1985, 8). Similarly, Zagzebski thinks that insofar as an epistemic subject is responsible for the acquisition of her true belief that *p*, the epistemic subject knows that *p*: "the knower *merits* the truth rather than merely acquiring the truth" (2012, 118).

3.3 (P3) - Properties with normative conditions cannot be natural kinds

In the previous section, the epistemic normativist's commitment to the claim that knowledge ascriptions are expressive, and not merely descriptive, was presented. Williams argues that natural kind terms are only deployed in descriptive claims, and as such, knowledge cannot be a natural kind. When employing a natural-kind concept one is making a descriptive statement about what "must, may, or cannot *happen*" (2015, 252; emphasis added). The use of a natural-kind concept is only appropriate when one is describing a state of affairs that must, may, or cannot *take place*. In contrast, social or normative concepts are used to describe states of affairs that ought, may permissibly, or ought not to occur. The application of social or institutional concepts is governed by rules regarding what "must, may or may not *be done*, which can of course be violated" (Williams 2015, 252; emphasis added).

Returning to our examples of gold and chess will be helpful in clarifying this distinction. While playing a game of chess, rule-following players may not move their rook diagonally across the board. If a player moves their rook diagonally across the board within the context of a game, they are in violation of the game's rules. Such a move may not be permissibly done during gameplay. It is not that rooks *never do* move diagonally across chess boards. They likely do with some frequency (for instance, when new players are learning the game). Such a move *can happen* but *may not* as a legitimate move in a chess game. In contrast, the property of being gold only and necessarily obtains when an object is composed of atoms with seventy-nine protons. It is not that a gold object composed of atoms with a different number of protons could occur, but be in violation of certain norms, standards, or rules. Rather, an object never has the property of being gold unless it is composed of atoms with seventy-nine protons.

The epistemic normativist claims that knowledge is, in part, modally governed in the same manner as the pieces of a chess game. We judge knowledge claims with reference to

epistemic rules, not natural laws that dictate whether or not a property is or is not obtaining. When ascribing knowledge to a mature adult human, we are not only judging whether it fits the necessary conditions such as those an object must meet to be an instance of an object of a particular natural kind (like gold or water), we are also judging whether the epistemic agent acted "in light of" certain norms. In virtue of having a justified true belief a mature human knower *may permissibly* assert and infer using that belief.[14] However, one's epistemic state does not limit or constrain whether one can in fact assert and infer using a particular belief.[15] What follows from a subject's justified true belief that *p* (and the property that one is a knower with respect to *p*) is not just a descriptive fact. Insofar as claims regarding natural kinds are purely descriptive, knowledge is not a natural kind.

The normative condition on knowledge also makes the category of human knowledge mind-dependent in the robust manner of institutional and social kinds (as opposed to natural kinds). Just as one cannot participate in a chess match (i.e. abide by the norms of chess) without having grasped requisite concepts (the constraints on the movements of the different pieces, the objective of the game), one cannot abide by a defensibility norm without understanding the rules governing epistemic justification. A child too young to fully understand the rules of chess may pick up a piece and move it across the board in a way that is consistent with the rules and objective of the game. However, we would not want to say that this child's action is an instance of gameplay. Legitimate skill in the game of chess requires that one's actions not *merely conform* to the rules and objective of the game, but that one's

[14] Williams: "The familiar idea that knowledge is *justified* true belief is naturally understood in deontic terms. To be justified in believing that *p* is to be *entitled* (in the light of epistemic standards) to use one's commitment to *p* in inferences (theoretical and practical), to inform others that *p*, and so on. Epistemic *permissions* are earned by fulfilling epistemic *obligations*" (Williams 2015, 253).

[15] One could use one's unjustified belief that *p* in inferences and inform others that *p*, all the while being unjustified in believing that *p*. Similarly, one could fail to use one's justified belief that *p* in inferences, and not inform other that *p*.

actions *intentionally and purposively conform* to the rules and objective of the game. A subject's purposive conformity to rules and objectives requires that they understand the rules and objective. If knowledge claims track epistemic entitlements that result from one's having fulfilled certain epistemic obligations, real epistemic authority requires not only that one's belief conform to standards of reliability. It also requires that epistemic subjects have the capacity to demonstrate (should the need arise) that their epistemic conduct assured that this standard of truth-reliability was met. This capacity to defend one's belief against legitimate inquiries requires understanding, or having a concept of, appropriate justification.[16] In order to abide by a normative defensibility condition, subjects need to have an understanding and a concept of justification – a concept of knowledge. A natural kind does not depend for its existence on our possessing certain concepts (Williams 2015, 251).[17] Human knowledge does and therefore, the epistemic normativists concludes, it is not a natural kind.

4. The Current Naturalist Reply

Epistemic naturalists have objected to epistemic normativists' argument by targeting (P3): "If knowledge has a normative condition, then knowledge is not a natural kind."[18] Kornblith (2011) claims that epistemic naturalism cannot be dismissed purely on the grounds that knowledge attributions track the fulfillment of standards or norms,

[16] One might take note here of the qualifications "should the need arise" and "legitimate". Williams does not think that all inquiries into epistemic responsibility are legitimate (and as such, not all failures to respond to inquiries defeat knowledge claims).

[17] To clarify, there are two manners in which the conclusion that the category of knowledge is mind-dependent (and therefore, not a natural kind) may be articulated. Insofar as being ascribed human knowledge requires that an epistemic subject be accountable to, and thereby understand, norms of epistemic responsibility, the property of being a human knower is mind-dependent. Similarly, insofar as knowledge is in part constituted by a corresponding dimension of epistemic responsibility that places an accountability condition on knowledge the content of which is *stipulated* by human practice, knowledge is mind-dependent.

[18] Indeed, Kornblith (2011) describes the target of his objection (to be outlined in this section) in the following manner:

> Normative standards and ideals are not items in the world that we discover. A standard or ideal is a paradigm case of something that we project upon the world. But this is just to say that, insofar as knowledge is a normative category, it is not a natural kind. (Kornlith 2011, 48)

because some standards are found in the natural world independent of human interest. His argument is as follows.

Not all norms are like the rules of chess. Cultures can codify non-anthropogenic norms in social practice. The systematization of a norm in human social practice is not a reason to think the norm is irreducible to features of the natural world that are independent of particular human interests. If there are reasons a standard should be valued independent of the norm-governed social practice in which it is embedded, we need not think that norm's existence is only explicable in reference to the anthropogenic or institutional practice in question. There are social practices and institutions surrounding health, Kornblith offers by way of example, but the standards of health are not ultimately determined by the aims of the social practice. Health is a normative concept insofar as being healthy "involves being in a certain sort of desirable state; if someone is unhealthy, there is something wrong with that person" (Kornblith 2011, 41). Although normative, health is not ultimately a standard that particular human communities project on the world. Rather, the norms of health reduce to the proper working of organic, natural functional kinds. The standards of proper functioning for each anatomical organ are not decided by human communities with reference to the community's particular interests. Organic functional kinds are the result of natural selection and as such serve a purpose that is independent of any particular human community's interests. Organic functional kinds are the result of, and serve, a species-wide interest: survival. The standards of proper function are determined by the conditions that need to be met in order for organs to perform the functions for which they were selected:

> The standards for proper functioning are not a matter of meeting standards that we project on the world as a result of our tastes or parochial interests. They are, instead, standards that flow from the very nature of the function that the organ was selected to perform. (Kornblith 2011, 48)

The notion of proper functioning picked out by "health" does not rely on human understanding or stipulation for its meaning. A heart riddled with bad cholesterol is unhealthy independent of whether social practices regarding the standards of health deem it so. Of course, Kornblith notes, institutional medical standards of health have changed through time. The current medical criteria for judging heart health is significantly different from the standards governing such judgements thirty years ago. However, those changes reflect increased understanding of objective features of the world, not merely our arbitrary or "parochial" interests (Kornblith 2011, 48). Changes in medical standards of heart health that resulted from the discovery of three major types of cholesterol did not track a change in the world independent of human understanding. Hearts were not *in fact* healthier prior to the discovery of different kinds of cholesterol. Identifying different types of lipid molecules merely enabled medical professionals to more accurately represent the state of a patient's physical well-being. This suggests that although our concept of health is in part constituted by norms and standards, we need not think the category of health is an institutional, as opposed to a natural, kind. The category of health is ultimately grounded in standards of proper functioning that are determined independent of human understanding or interest.[19]

Kornblith argues that a similar conclusion can be drawn in the case of knowledge. Even if one thinks that there is a normative dimension to human knowledge, one need not be

[19] Note that with this analysis, we are now able to see that ultimately norms of health are explicable in non-normative descriptive terms (as opposed to normative health-related terms). Consider the following formulations. To emphasize the distinction between them, the normative health-related terms in the former have been underlined.

> *Articulated in normative health-related terms:* To be deemed <u>healthy</u>, patients <u>should</u> be <u>responsible</u> and moderate the about of dairy and red meat in their diet so as to maintain an <u>optimal</u> LDL (low-density lipid) cholesterol level.
>
> *Articulated in non-normative, descriptive terms*: The human heart is most likely to perform the function for which it was selected (pumping blood) and enable the survival of the organism whose heart it is, when levels of LDL (low-density lipid) cholesterol are below 100 ml/DL.

committed to the conclusion that the category of knowledge is not a natural kind. Epistemic norms governing epistemic conduct are akin to the standards of health that govern proper medical care. Human "cognitive equipment," responsible for the production of true beliefs, is just as much an organic functional kind as the human heart (Kornblith 2011, 48). Like other organic functional kinds, human cognitive equipment should be evaluated by gauging its success at achieving the function for which it was selected. Again, this function is not determined by the interests of a particular human community. Rather, the proper functioning of human cognitive equipment is the result of, and serves, a species-wide interest. Both epistemic and health-related standards are not parochial or purely institutional, but rather natural phenomena insofar as there are reasons to abide by them independent of the ends of their associated institutional practices. There are species-wide (robustly mind-independent) reasons to value health independent of the standards of health stipulated by the medical community. Some level of health is necessary for survival. Similarly, there are species-wide reasons to want true beliefs. True beliefs help an epistemic subject successfully navigate her world and pursue her projects, regardless of her community's epistemic norms and the particularities of her goal. In other words, there are reasons independent of the particular epistemic values and norms of human epistemic communities – i.e. robustly mind-independent reasons – to value true belief:

> True beliefs play a crucial role in the production of successful goal satisfying behavior, and there is evolutionary pressure to select for cognitive processes that are instrumental in producing behavior that satisfies biologically given needs. (Kornblith 2011, 45)

Even the "loner" who does not participate in the epistemic community, Kornblith argues, "might well come to the belief that not just anything goes when it comes to belief formation" because reliably formed beliefs are valuable for any member of the human species (and some

non-human species), not just members of particular epistemic communities (2002, 95). Reliably formed beliefs help us, and some non-human animals, survive and thrive in our given environments. At the most fundamental level, knowing the location of food, shelter, danger, and safety is central to survival. More broadly, reliable beliefs aid in the successful pursuit of projects regardless of what they projects are, and the particular epistemic norms of the society in which one finds oneself. As such, epistemic standards are not just the result of the interests of particular epistemic communities. Epistemic standards are ultimately explicable in non-epistemic terms used to describe of the proper functioning of an organic function kind (in this case, human cognitive equipment) that was selected for in light of the evolutionary interests of a species in its natural environment.[20]

4.1 A Rebuttal on Behalf of the Epistemic Normativist: it doesn't follow that all epistemic norms are natural

Epistemic normativists may remain unconvinced by this naturalist response as it arguably fails to show that norms of epistemic responsibility *in particular* are natural norms. Why think that the norms governing epistemic responsibility *in particular* are like those governing health and not like those governing the movement of chess pieces? The naturalist reply, sketched above, is that insofar as there are non-anthropogenic reasons to want true

---

[20] Note the extent to which this analysis parallels that of health. As before (footnote 19), we are now able to see that ultimately epistemic norms can be explained in non-epistemic, non-normative terms. Consider the following:

> *Articulated in normative, epistemic terms:* To be deemed knowledgeable, epistemic subjects should be responsible in the acquisition of beliefs. They should be capable of justifying their beliefs. They should be able to demonstrate their responsibility through the provision of reasons in defense of their beliefs.
>
> *Articulated in non-normative, non-epistemic terms:* Human cognitive equipment is most likely to perform the function for which it was selected (the formation of true beliefs) and enable the survival of the organism whose cognitive equipment it is when the conditions necessary for reliable belief-forming processes obtain.

Note that this latter formulation "*articulated in non-normative, non-epistemic terms"*, uses exclusively vocabulary Goldman takes to be non-epistemic: "In general, (purely) doxastic, metaphysical, modal, semantic, or syntactic expressions are not epistemic (1979, 2).

beliefs, all norms governing knowledge reduce to natural standards of reliability. An epistemic normativist might find this reply a plausible explanation of some epistemic norms, but not *all* epistemic norms. The epistemic normativist could concede that social epistemic practices are, at least in part, grounded in a natural norm while maintaining that our entire epistemic institution cannot be so explained.

To clarify, consider again the following theses: (a) health is a natural norm and, (b) medicine is a social institution with standards and practices that can be explained in terms of that natural norm. It seems one could quite plausible adopt the following weaker version of (b): *many* (if not most) medical standards and practices are explicable in terms of (a), but there can be socially-teleological elements of medical practices and norms that are inexplicable solely in terms of (a). For example, the question "is this person healthy enough to play in Saturday's game?" asks if an individual is functioning properly enough to be able to participate in a socially-constructed practice. "Healthy enough to play in Saturday's game" is a standard with its roots in a natural norm regarding the proper operating of organic functional kinds, but sets that norm relative to particular, social ends. "Healthy enough to play in Saturday's game" picks out no single socially-independent standard that obtains in the natural world. Rather, this standard is also a function of a social practice: Saturday's game.

A similar concern can be raised regarding the relationship between reliability as a natural norm, and social practices governing responsible epistemic conduct. Williams, for example, does grant that reliability is a natural standard that explains all social epistemic norms.[21] Indeed, he claims that norms of epistemic responsibility ought to be updated in light

[21] Williams grants the following: "Epistemic norms are explained by reliability. Obviously, we have an interest in having true beliefs, for having true beliefs is important for achieving our purposes, whatever they happen to be" (2004, 207).

of what we discover regarding their truth-conductivity.[22] However, this concession does not commit him, or any epistemic normativist, to understanding epistemic norms as completely derivative of, or reducible to, standards of reliability. Williams claims as much in describing reliability as "a norm that we are not only responsible to but, in certain applications, responsible *for*" (Williams 2015, 286; underline mine). His evidence in support of this claim are those cases in which human understanding and interest *determines* the appropriate methods and standards of inquiry. Note the way in which standards of reliability are context dependent and variable, such that one and the same kind of epistemic conduct is responsible in one context and irresponsible in another. The testimony of a stranger may be an epistemically responsible way to find out the time when en route to a non-urgent appointment, but epistemically irresponsible conduct when collecting data within the confines of a scientific experiment. The reliability of the testimony of strangers is not necessarily different in each of these contexts. Rather, human interests and purposes are different in the two situations, which renders processes of identical reliability responsible in one situation, and irresponsible in another. The epistemic normativist claims that this is evidence that human knowledge is a function of both (mind-independent) standards of reliability and (mind-dependent) purposes of the inquiry in question.

In clarifying this claim Williams offers the following example. Advancements in particle physics have prompted changes in "the standard for "detecting" a particle" (Williams 2015, 267-268). The standard for proper detection "has moved from three to five sigma, the standard in effect when the discovery of the Higgs boson was announced" because ""discoveries" at three sigma – itself a high standard – have sometimes turned out to be

[22] Williams explicitly makes this claim: "I think that our ideas about justification change and should change in accordance with our interest in improving our reliability" (2004, 205).

statistical blips" (Williams 2015, 267-8). This demonstrates, Williams argues, that reliability standards are not always established in reference to natural standards, but rather particular human uses and purposes. Human understanding is responsible for setting the standard at five sigma because such a standard establishes reliability *for the purpose of* yielding data of interest and use to particle physicists. The norm of reliability in the context of mature human knowledge is, Williams argues, distinct from the norms of reliability governing non-human animal belief insofar as the notion of reliability relevant to human epistemic practices is necessarily teleological. In the context of mature human knowledge, "reliability itself becomes reliability for particular purposes" (Williams 2015, 267). The purpose that sets the standard of reliability in human epistemic contexts is non-natural, Williams argues, because it is not dependent on features of the environment (objective features of nature), but rather human understanding. In this way, epistemic norms are not, as Kornblith would describe, "merely a reflection of the social recognition of pre-existing normative demands on belief, demands that are not social in origin" (Kornblith 2002, 93). This is because norms don't track reliability *simpliciter*, but rather reliability *for the purposes of* the epistemic community.[23] Given this social, teleological dimension, the category of human knowledge cannot be articulated solely in non-epistemic, non-normative terms. The establishment of standards of reliability relative to an epistemic community's parochial aims requires acknowledging that epistemic community's epistemic concepts. Being a knower by the lights of the particle physicist's community requires making reference to the community's (five sigma) concept of reliability, which can only be articulated in epistemic and normative terms. Knowledge needs an institutional dimension, a

[23] This institutional, teleological analysis of knowledge can also be found in Craig's *Knowledge and the State of Nature:* "Knowledge is not a given phenomenon, but something that we delineate by operating with a concept which we create in answer to certain needs, or in pursuit of certain ideals" (1990, 3).

necessarily normative epistemically-articulated defensibility condition, to account for the way in which the standards of reliability are set relative to particular, parochial aims.

5. A New Naturalist Reply: defensibility as species-specific adaptation to the privileged human environment

Putting into sharper focus the debate between the epistemic naturalist and normativist helps to identify precisely what the former must claim to respond to the latter's concern. A successful rebuttal requires a naturalist to argue that the *anthropogenic purposes* used to establish standards of reliability in human epistemic practice are derivative of natural norms. The naturalist must argue the following: the human epistemic practice of establishing norms of reliability relative to parochial purposes is part and parcel of standards regarding the proper functioning of human cognition. Standards regarding the proper functioning of human cognition are, as Kornblith has argued, "not a projection of human parochial interests and concerns" (Kornblith 2011, 48). Therefore, behaviors that fall under the umbrella of proper function are, in an important sense, independent of parochial interests as well.

Williams argues that this kind of explanation is not available to epistemic naturalists insofar as anthropogenic norms established for particular purposes are not features of the natural world, but rather are dependent on human understanding. Consider again the standards of inquiry in particle physics. The standard "five sigma" is established because it helps empirical scientists gather and share their data given the current state of human understanding (the extent to which we comprehend the laws governing the natural world and the tools we have for investigating it). In this section, I will argue that contemporary work in evolutionary and social psychology provide resources for understanding this teleological element of human epistemic practice as following from the proper functioning of human cognitive capabilities in their natural environment. I will conclude that the capacity for setting

purposive standards of reliability is, in a meaningful sense, mind-independent insofar it follows from a species-specific adaptation.

I will begin by briefly sketching the Darwinian commitments of an epistemic naturalist, like Kornblith. Reliably formed beliefs (or more generally information-bearing states) have natural normative status because there are socially-independent reasons to value them.[24] True beliefs enable humans, and other creatures that represent both "needs together with features of the environment," to "deal effectively" with their environment (Kornblith 2002, 37-38). A species-wide capacity to form true beliefs about the environment aids that species' survival. The capacity for knowing (or at least representing) what is dangerous, where dangerous things are, what is life-sustaining food, and where to find it is adaptively advantageous. This epistemic capacity is valuable independent of its social role. Furthermore, fitting and well-adapted belief-forming processes are those which enable a species to navigate their particular, i.e. privileged, environment well (Kornblith 2002, 65). For example, echolocation in nocturnal bats is a fitting belief-forming process, well-suited to nocturnal bats' natural environment. Similarly, it is apt that many birds of prey have exceptional eyesight. The natural environment of birds of prey is one in which they need to identify sources of food from great heights. The features of a

[24] This is a brief overview of Kornblith's argument for one of his central claims: "Knowledge is an ecological kind: it has to do with the fit between an organism and its environment" (2002, 65). It will do for the purposes of this paper but is an overview of more complex and detailed commitments Kornblith draws from engaging with cognitive ethology. See Kornblith (2002), in particular Chapter 2.3.

Importantly, the outline here is consistent with Williams' understanding of Kornblith's "perspective of cognitive ethology" (Williams, 2004, 203). Williams' take Kornblith to be committed to the following:

> 1. An animal's cognitive abilities are seen as the product of natural selection. The best explanation of the animal's cognitive capacities is that they were selected for.
> 2. Behavior that contributes to fitness makes certain informational demands on the animal, and the animal's cognitive capacities were selected for their ability to play this role. (Williams 2004, 203)

Note that these claims are not quite disputed by Williams. Rather, he thinks that they are not fully descriptive of human epistemic practice.

species' natural environment put evolutionary pressure on the species' cognition to adapt abilities that will aid in the successful navigation of that particular environment.

What is the natural, human environment? Most species of non-human animals occupy a particular geographical region with distinctive features. The human species, however, occupies various and diverse parts of the world. Recent work in evolutionary, behavioral and comparative psychology suggests that human capacity for metacognition, reflective deliberation on one's beliefs, and thinking generally, ought to be understood as adapted in response to, and for, the human sociocultural environment. In *A Natural History of Human Thinking*, developmental and comparative psychologist Michael Tomasello attempts to "reconstruct the evolutionary origins" of "unique human objective-reflective-normative" thinking. While it is fitting to understand the geographically-determined features of the natural world as primarily (or predominantly) responsible for the adaptation of particular cognitive capabilities in many non-human animals, the same is not so easily said of humans. While human cognition was undoubtedly significantly shaped by non-social features of the environment, such features cannot explain the evolution of human cognition in its entirety. Human cognitive capabilities also meaningfully adapted to the interpersonal, or community-oriented features of the species' natural environment. Tomasello develops this conclusion by arguing that human survival in various and diverse geographical locations was (and is) possible because human cognitive capacities enable our species to collaboratively develop location-specific practices and strategies.

> Unlike other great apes, who all live in the general vicinity of the equator, modern humans have migrated all over the globe. They have done this not as individuals but as cultural groups; in none of their local habitats could a modern human individual survive for very long on his own. Instead, in each specific environment, modern human cultural groups have developed collectively a set of specialized and cognitively complex cultural practices to accommodate to local conditions, from seal hunting and igloo building to tuber gathering and bow-and-arrow making – not to mention science and mathematics. (Tomasello 2014, 120).

Tomasello argues that the evolution of human cognition ought to be understood using the "*shared intentionality hypothesis*" (Tomasello 2014, 4). This hypothesis claims that the distinctive features of human thinking, "its process of presentation, inference, and self-monitoring" are "adaptations for dealing with problems of social coordination, specifically, problems presented by individuals' attempts to collaborate and communicate with others (to co-operate with others)" (Tomasello 2014, 4). Tomasello identifies two major evolutionary pressures on the human species to develop cognitive capabilities necessary for social cooperation over and above the "ad hoc collaborative foraging characteristic of early humans" (Tomasello 2014, 82). Competition with other groups as well as increasing population sizes made it such that the ability to identify with one's group was to members' advantage: "the subterranean effect of this wave of group-mindedness and conformity, as it were, was new and culturally collective forms of cognitive representation, inference, and self-monitoring for use in thinking" (Tomasello 2014, 121). Collaboration requires participants to have metacognitive capacities, concepts of a group perspective and group standards that inform their behavior. Abiding by group conventions requires that members have concepts of what those conventions are, and the belief that others in their group expect them to abide by such concepts: "the group has its perspective and evaluations and I accept them; indeed, I myself help to constitute the group's perspective and evaluations, even if the target is myself" (Tomasello 2014, 92). Individual members of the group understand what features of their experience are shared by their group members, the norms the group collectively uses to evaluate that shared experience, and how those norms ought to constrain their own behavior.

Here, we have the anthropogenic standards of reliability, a condition of defensibility, explained through an empirical, naturalistic lens. I argue that the claims the epistemic normativist uses to support her conclusion that the category of knowledge is not a natural

kind when viewed through this lens no longer favors her conclusion. Rather, these same considerations suggest that the category of (even mature human) knowledge is a natural kind.

Recall that I presented the epistemic normativist as citing two primary sources of evidence for (P2), the claim that "there is a normative condition on knowledge because knowledge claims are not purely descriptive." The first source of evidence concerned the way in which we have epistemic interest in knowledge that is separate from our interest in true belief. We value both reliability and trust. That human knowledge expresses both values (even often, if not always) is fitting against the backdrop of collaboration. To work with others, one needs to know whether the testimony of one's interlocutors is reliable. I will not know to what use to put the contributions of my interlocutors in developing a practice to aid in the survival of our community if I know nothing about my interlocutor's reliability. Establishing group standards of reliability, and then using knowledge attributions to signal the fulfillment of such standards solves this problem. Importantly, that knowledge claims do so is not because of any particular parochial, human interest. This capacity and practice developed independent of any particular mind, or particular community of minds. Although responsive to a sociocultural environment, the practice is not fundamentally anthropogenic. Understood as an adaptation that helps a member of the human species navigate its privileged environment, behavior that broadly satisfies the epistemic normativist's defensibility condition is widely categorized as such independent of the stipulated norms of any particular epistemic community. Specific human interests are not responsible for the formation of such epistemic practices; rather species-specific interests are responsible.

Consider the epistemic normativist's rebuttal to Kornblith's objection (Section 4.1), i.e. the claim that insofar as human epistemic standards are established relative to human understanding, the category of human knowledge cannot be a natural kind. Insofar as

community-relative norms of epistemic responsibility are dependent on a particular social group for their existence, the epistemic normativist thinks that the category of human knowledge cannot be a natural kind. However, as discussed above, collaboration requires that there be standards of reliability useful for the purpose of the collaboration: standards that can be appreciated by all participants given their particular breadth and depth of understanding. Establishing standards of reliability, of inquiry, within a group, is necessary for collaboration. The particular project on which humans are collaborating may dictate what the relevant standards of reliability are. Consider again Williams' example of the particle physicists. Collaboration between particle physicists is made possible by the stipulation of project-specific norms of inquiry. In order for a particle physicist to understand how her colleagues' data relates to her own research, she has to know that both their and her own data was collected in a manner consistent with the shared perspective and evaluations of particle physicists in particular. However, the capacity for setting these project-relative standards is selected for the fulfillment of species-wide interest and need. As such, the capacity for epistemic behavior that satisfies the epistemic normativist's defensibility condition – the capacity requisite for being a knower by the lights of the epistemic normativist – is not dependent on the projection of the interests of particular human community. Therefore, the epistemic interests and behavior the epistemic normativist cites as evidence of a defensibility condition, when viewed through the naturalist's lens, can be explained in non-normative, non-epistemic terms.[25] The ability to set relative, teleological standards of reliability relevant to the interests of one's community is a

[25] To make this explicit (and to parallel the analysis given in Section 4, footnotes 19 and 20, consider the following formulation:

> *Articulated in non-normative, non-epistemic terms*: Human cognitive equipment is most likely to perform the function for which it was selected (the formation of true beliefs) and enable the survival of the organism whose cognitive equipment it is, when metacognitive capacities are deployed in dialogical, collaborative contexts.

necessary result of the proper functioning of cognitive equipment that was selected for given a species-wide interest.

6. The defensibility condition is modally governed in the manner of natural kinds.

At this juncture, the epistemic normativist may be satisfied that the above considerations give an explanation of the first source of evidence for (P2) – our epistemic interest in trust – in non-normative, non-epistemic terms. There are non-epistemic, non-normative reasons to want knowledge ascriptions to express trust: signaling and labeling trust is essential for collaboration which enables humans to navigate their species-specific environment. Furthermore, the epistemic normativist may be satisfied that the rebuttal to Kornblith's objection has been sufficiently refuted. The purposive element of institutional standards of reliability follows from an adaptive, natural capacity that serves species-wide (as opposed to institutional) interests. However, the epistemic normativist may remain skeptical insofar as this naturalistic lens has not yet accounted for the second source of evidence for (P2): that knowledge ascriptions are modally governed in the manner of institutional, not natural, kinds. If knowledge is a natural kind, the epistemic normativist may question: Why is it the case that epistemic subjects are ascribed the property of knowing in a manner analogous to a chess piece being ascribed the property of being in check? Why is it the case that epistemic subjects are ascribed the property of knowing in a manner dis-analogous to a piece of metal being ascribed the property of being gold?

While it is true that ascriptions of knowledge track normative entitlements of trust as well as adherence to standards of reliability, behavior that broadly satisfies the epistemic normativist's defensibility condition is prompted with law-like consistency by dialogical contexts. Recent work from psychologists Hugo Mercier and Dan Sperber (2011, 2017) suggests that epistemic practices associated with the epistemic normativist's defensibility

condition fits the description of a natural, non-institutional phenomenon insofar as the occurrence of the relevant behavior (the giving of, and asking for, reasons) nomologically *occurs* in precisely the collaborative contexts definitive of the species' privileged environment discussed above. Just as there is necessarily and nomologically an instance of gold when there is metal composed of atoms with 79 protons, there is necessarily and nomologically an instance of epistemic behavior associated with the epistemic normativist's defensibility condition in dialogic, collaborative contexts.

In a manner consistent with Tomasello's work, Mercier and Sperber posit an "interactionist" theory of reasoning: a teleological analysis of what reasoning *is* based on what it is *for* (M & S 2017, 9). This theory emerges from the examination of those instances in which reasoning functions optimally through the lens of evolutionary psychology. Their conclusion on examining those cases of optimal function is that reasoning works well in "social, and more specifically dialogic" contexts (2011, 247). If it is the case that human cognition adapted in response to the human species' sociocultural environment, it is fitting for deliberative and justificatory capacities to have evolved specifically for the navigation of this environment. Mercier and Sperber argue that this is precisely the case, and that reasons are "social constructs" developed in response to the evolutionary pressures humans faced in virtue of the fundamentally social environments they had to navigate (2017, 127). Informing this conclusion is empirical evidence that suggests our reasons for belief are primarily retrospective justifications developed in response to anticipating and engaging in dialogic contexts. The data Mercier and Sperber analyze and discuss in defending this claim shows that human subjects are better at both evaluating and producing arguments in dialogic contexts than when

reasoning on their own.[26] Human subjects are better able to identify argument structure and anticipate counterarguments in dialogic contexts, suggesting that the cognitive capacities responsible for reasoning were adapted in response the evolutionary pressures of dialogic contexts (M & S 2011, 61).

In describing reasons as "social constructs", Mercier and Sperber are not claiming that the act of producing reasons in support of belief is not a natural phenomenon. Rather, they argue that data shows the contents of reasons are constructed, or formulated, for "social consumption" (M & S 2017, 127). On evaluation of the empirical research, Mercier and Sperber conclude that the reasons we cite for our beliefs are geared towards the epistemic satisfaction of our community, and in fact have little to do with the actual basis of our beliefs (M & S 2017, 127). The contents of our articulated reasons are formulated by "distorting and simplifying our understanding of mental states and of their causal role and by injecting into it a strong dose of normativity" (M & S 2017, 127). The provision of reasons for belief, Mercier and Sperber claim, track what we think will be convincing to our epistemic community, not the actual inferential processes employed in their formation or evaluation. Supporting this conclusion is research that demonstrates that the reasons given for belief are often confabulated (M & S 2017, 114-115).

These findings, I argue, are evidence that there is an important sense in which human epistemic practices requisite for satisfying the epistemic normativist's defensibility requirement are modally governed in the manner of natural kinds. Behavior associated with the epistemic normativist's defensibility condition is governed by purely descriptive facts about the world: the presence and demands of collaborative dialogical contexts. While knowledge ascriptions

[26] Kornblith uses evidence that suggests reflective reasoning is quite poor in isolation to refute non-naturalist claims that metacognition, or reflective deliberation, is valuable insofar as it increases reliability. See Kornblith 2012, Chapter 1.3.

express epistemic entitlements, Mercier and Sperber's interpretation of the empirical research suggests that an epistemic subject's provision of reasons in support her beliefs tracks with law-like consistency the occurrence of dialogical collaborative contexts. Their analysis suggests that the deployment of metacognition, reflective deliberation, or the assertion of reliability knowledge is a phenomenon that occurs independent of parochial human interests insofar as it functions optimally with nomological uniformity in response to the presence of dialogic contexts.[27] It is not merely that instances of reflective deliberation in dialogic contexts yield certain kinds of normative judgments, for example, that it was epistemically responsible for reflective deliberation to occur. Rather, human reflective deliberation, or the production of reasons, is nomologically necessitated by an epistemic subject's participation in collaboration, a context in which epistemic authority requires that they defend the reliability of their knowledge claims.

Mercier and Sperber's interactionist thesis accounts for optimal functioning of the capacity to produce reasons in dialogic contexts by arguing that such behavior is valuable to the epistemic projects of groups. They claim that individual reflective reasoning performed in isolation is poor because it is primarily a retrospective justificatory process (2011, 66, 129; 2017, 127). Insofar as the contents of reasons are crafted for social consumption, the process of consciously evaluating one's reasons fails to be reliable when one is evaluating, i.e. consuming, by oneself. The central piece of data for this argument is the breadth of research that has consistently documented the "myside bias": human reasoning's "preference for confirmation" (M & S 2017, 218). The human cognitive capacity to produce reasons is motivated not to identify the truth, but to persuade others. As such, we systematically find

[27] By "the deployment of metacognition, reflective deliberation, or the assertion of reliability knowledge" I mean those epistemic behaviors that enable an epistemic subject to fulfill a defensibility condition, i.e. behaviors that make the epistemic subject capable of citing reasons in defense of her belief.

reasons "for our ideas and against the ideas we oppose" (M & S 2017, 218). The data shows, Mercier and Sperber claim, that a subject's reasoning will "always take [their] side" (M & S 2017, 218). While this process is epistemically detrimental for reasoners working in isolation, in dialogical contexts it serves as a "division of cognitive labor" (2017, 221). If interlocutors submit evidence and arguments in favor of their divergent positions for group evaluation, and the group then assesses the relative merits of those pooled resources, the group is more likely to arrive at the correct answer to their question, or solution to their problem:

> The myside bias makes reasoners focus on just one side of the issue rather than having to figure out on their own how to adopt everyone's perspective. Laziness lets reason stop looking for better reasons when it has found an acceptable one. The interlocutor, if not convinced, will look for a counter argument, helping the speaker to produce more pointed reasons. By using bias and laziness to its advantage, the exchange of reasons offers an elegant, cost-effective way to solve a disagreement. (M & S 2017, 236).

By attempting to fulfill norms of epistemic responsibility, individual epistemic subjects contribute to a uniquely human method of belief improvement and correction. In articulating reasons for beliefs, human epistemic subjects enable their epistemic community to resolve disagreements and help update the norms of epistemic responsibility so that they better align with standards of reliability. In essence, responsible epistemic conduct is ultimately reducible to our interest in having reliable beliefs, i.e. explicable in non-epistemic terms. In a sociocultural environment, the processes by which reliable beliefs and belief-forming mechanisms are cultivated requires the employment of metacognitive capabilities adapted for such dialogic contexts. Human epistemic conduct within such contexts is natural insofar as it is a species-specific adaptation for forming reliable beliefs given the particularities of the privileged human environment.

7. Is the defensibility condition still viable?

Before concluding, I want to make clear the scope of this paper's project. The arguments developed here contend that while the epistemic normativist's defensibility condition captures an interesting and important part of human epistemic practice, there is no reason to think this practice (if it is partly constitutive of human knowledge) makes it the case that either (a) human knowledge is not a natural phenomenon or (b), the category of human knowledge is not a natural kind. In considering these arguments, one might be convinced that the rejection of both (a) and (b) are plausible but remain puzzled as to the viability of an independent defensibility condition on knowledge.[28] By way of conclusion, I will discuss two possible ways of proceeding: a naturalized epistemic normativist route, and an austere reliablist route. I will begin with the former.

If convinced by the arguments presented above, a naturalized epistemic normativist might decide to concede that (a) and (b) are false while maintaining the defensibility condition. The proponent of this position would claim that while defensibility is ultimately explicable in non-normative and non-epistemic terms, it remains an independent condition on the category of human knowledge. The naturalized epistemic normativist might motivate this position by separately arguing that the category of human knowledge is distinct from the category of non-human animal knowledge (or more conservatively, non-human animal information-being states). Stipulating a naturalized independent reliability condition would maintain such a distinction while granting the conclusions argued for above. The naturalized epistemic normativist might also argue for the independence of the defensibility condition on the same grounds Brandom employs in claiming that reliabilism has a serious "Conceptual Blindspot"

[28] By "independent", I just mean separate and distinct from the reliability condition to which epistemic normativists were described as antecedently agreeing.

(Brandom 1998, 381). Brandom argues that in order for reliably formed beliefs to serve the purpose of aiding in the successful navigation of epistemic subjects' environment, the epistemic community needs to employ a concept of defensibility to discriminate between reliable and unreliable indication (i.e. representation of the external world). Although members of a community without a concept of justification (i.e. a concept of "giving reasons for beliefs") may be "measuring instruments," "they cannot treat themselves or each other as" reliable indicators in the absence of this concept of epistemic assessment (Brandom 1998, 379). Underlying Brandom's argument is the assumption that an epistemic subject need take herself to be an instrument of reliability to use her reliably formed beliefs in the successful navigation of her environment: "the very notion of a correlation between the states of an instrument and the states that it is a candidate for measuring is unintelligible apart from the assessments of reliability" (Brandom 1998, 379). A concept of justification is required for taking both oneself and one's peers to be reliable indicators, which is requisite for collaboration. The Brandomian might argue that while an independent defensibility condition on knowledge does not render the category an institutional or social kind, it is necessary insofar as it enables the successful navigation of the species' environment cited above as evidence for the conclusion that such epistemic practices are ultimately explicable in non-normative, non-epistemic terms.

In considering this (naturalized epistemic normativist) route, one might be skeptical of the assumption that an epistemic subject needs to take her belief to be reliable to use it in the successful navigation of her environment. Furthermore, one might be concerned that positing a defensibility condition that is reducible to non-epistemic, non-normative reliability-terms is redundant insofar as the epistemic normativist antecedently posits an externalist reliability condition. Those who pursue this line of criticism will likely be more partial to the

second, austere reliabilist route. The austere reliabilist will take the arguments above to vindicate their original position, despite dispelling the false dichotomy described at the outset of this paper. The practice of giving and asking for reasons, the austere reliabilist might claim, does have epistemic significance – just not as a constitutive feature of knowledge. The practice of giving and asking for reasons is of epistemic interest because it is part of species-specific reliable belief-forming processes. This line of thought is pursed by Jeremy Fantl (2015). In a direct reply to Williams (2015), Fantl argues that it does not follow from the expressive nature of knowledge ascriptions that responsibility is a constitutive feature of human knowledge or that human knowledge is different i*n kind* from non-human animal knowledge. Rather, human social practices regarding epistemic responsibility suggest that *human knowers* have a species-specific accountability feature.[29] Defensibility is not essential to human knowledge, but merely a supplemental feature of human epistemic practice (that describes epistemic conduct over and above that which is required for knowledge).

What I have shown is that regardless of the stance one takes on this question, there is no reason to think that the category human knowledge is not a natural kind. While it is outside the scope of this paper to fully adjudicate between these options, the considerations of this paper suggest a new framework for settling the question. In essence, epistemologists ought to consider what role, if any, the species-specificity of a process plays in making epistemological and kind distinctions. Does the fact that defensibility is a species-specific adaptation for forming reliable beliefs in the privileged human environment render internalist justificatory capacities necessary for a distinctly human kind of knowledge?

[29] In short, Fantl argues that Williams' arguments do not suggest the following robust thesis: the simultaneous attribution of epistemic authority and epistemic accountability in human practice shows that human knowledge is *constituted* in part by an evaluative epistemic responsibility dimension. Rather, Fantl thinks that Williams' arguments are only evidence of "the weaker thesis – that human knowers are accountable in ways that animal knowers are not" (Fantl 2015, 272).

In answering this question, it is helpful to consider a case of a non-human species-specific reliable belief forming process adapted for a particular, privileged environment. For example, nocturnal bats need to successfully navigate an environment with little light. Many species of nocturnal bats have adapted an ability to echolocate, a perceptual mode that aids in their navigation of this environment. Such bats "see" in the dark by emitting sounds and listening to the associated echoes that bounce off solid objects in their surroundings. Just as the capacity for satisfying a defensibility condition is well-adapted for helping members of the human species navigate their natural sociocultural environment, echolocation helps members of nocturnal bat species navigate their natural nocturnal environment. Is an echolocating bat's information-bearing state formed by means of echolocation "special" insofar as there is an evolutionary match between the process by which it was formed and the distinctive feature of the creature's natural environment? Consider an individual member of an echolocating species of bat that cannot itself echolocate. Of this individual bat, would we want to say that it is incapable "bat knowledge"? This conclusion, I argue, does not aptly describe such a case. The bat in question is not capable of forming beliefs (or information-bearing states) about its environment in the species-specific manner developed in response to its natural environment. However, it is likely able to form reliable beliefs through other sensory modes. As such, it seems that the strongest claim one is justified in making is that a non-echolocating nocturnal bat is incapable of being the same kind of epistemic creature as the other (echolocating) members of its species — not that it is incapable of acquiring a distinct kind of knowledge.

My tentative suggestion is that a similar distinction could be made in the context of human epistemic practice. Practices surrounding the epistemic normativist's defensibility condition are not evidence that there is a distinct kind of human knowledge in part constituted by the norms associated with such practices. Rather, human social epistemic conduct that

requires the provision of reasons for belief is part of a "special" kind of belief forming process; "special" insofar it is an evolutionary adaptation designed for the privileged human environment. One could plausibly suggest that (mature adult) humans need to be capable of participating in social epistemic practices in order to be the kind of epistemic subject that members of the human species are. However, this makes the category of human knower "special", not the category of knowledge.

8. Conclusion

This paper has argued that the category of knowledge is a natural kind. Analysis of evolutionary and social psychological data was used to demonstrate that human epistemic practices associated with the epistemic normativist's defensibility condition can be explained as flowing from a natural human capacity adapted in response to the species' sociocultural environment. Furthermore, empirical data was used to support the claim that the occurrence of such epistemic practices is modally governed in the descriptive manner of natural kinds. In arguing that these findings are evidence for the conclusion that knowledge is a natural (given) phenomenon, my aim is to clarify a point of disagreement between epistemic normativists and naturalist, austere reliabilists. Importantly, the dichotomy introduced at the outset of this paper has been dismissed. My arguments have shown that a naturalist, austere reliabilist theory of knowledge is compatible with casting the capacity for conscious, reflective evaluation of beliefs in an important and significant epistemic role. Although it vindicates the naturalist epistemological research project and its methodologies, this paper also demonstrates that this research project is consistent with finding an important place for rational reflection in our theorizing of human epistemic practice.

CHAPTER 2

A TRULY, MADLY, DEEPLY SOCIAL THEORY OF EPISTEMIC JUSTIFICATION

1. Introduction

Social epistemologists have long criticized traditional epistemology for its myopic focus on individual epistemic subjects. They have taken issue with the idea that individuals are the central loci of epistemic evaluation (e.g., Gilbert 2004, Lackey, ed. 2015) and the idea that individuals are solely responsible for what they know (e.g., Lackey 2007, 2009). They have argued that, in traditional epistemology, insufficient attention has historically been given to the epistemic significance of testimony and other social forces (e.g., Longino 1990 and 2002, Goldman 1999, Goldberg 2010). However, even many social epistemologists largely follow in Descartes' individualist tradition when it comes to our ability to reason. The standard view is that reasoning is a private competence, and the social enters into the epistemological picture when we start exchanging the outputs of that private competence with one another. Our interlocutors are information sources, like thermometers, and our primary social epistemic task is to evaluate and respond appropriately to their reliability (just as we do with measuring instruments generally). However, new research concerning our ability to reason puts pressure on this approach. In *The Enigma of Reason*, Hugo Mercier and Dan Sperber develop and defend an "interactionist account" according to which reasoning is a social competence that yields epistemic benefits for individuals through social interaction with others (2017, 9). This paper explores how epistemologists should incorporate the interactionist account into a theory of doxastic justification. I argue that the epistemological consequence of Mercier and Sperber's position is a yet unexplored kind of social epistemic dependence on others, one not immediately characterizable in terms of epistemic reliance on interlocutors' reliability. Working on the process reliabilist's framework, I argue that in light of (1) the plausibility of historical theories of justification and (2) Mercier and Sperber's interactionist theory, we should posit the existence of extended interactive justification-conferring

processes. My central claim is that, in cases where beliefs are formed and sustained by dialogical deliberation, the relevant process that confers justification on a deliberative participant's belief doesn't occur solely within the cognition of that particular subject, but rather extends beyond it to include her interactive engagement with other deliberative participants. I argue that a significant consequence of this claim is that not all that is justification-conferring is evidential. As such, the foregoing analysis not only supports reconceiving the process reliabilist's notion of justification-conferring processes, but it also serves as an argument against evidentialism.

I start with argument in favor of historical accounts of justification. I then highlight that this argument makes perspicuous an explanatory desideratum for theories of doxastic justification. After, I highlight and present relevant parts of Mercier and Sperber's interactionist theory. I then argue that accepting Mercier and Sperber's theory and satisfying the explanatory desideratum for justification requires that process reliabilists accept extended interactive belief-forming processes. In the following section I turn to the project of demonstrating how this analysis can also be used to formulate an argument against evidentialism given that evidentialism can't satisfy the explanatory desideratum if Mercier and Sperber's interactionist account of reasoning is true.

2. Historical Accounts of Justification and the Explanatory Desideratum

In this section, I review the persuasive case in defense of the claim that doxastic justification has an important historical dimension. I then argue that this defense highlights a broader, more demanding desideratum any plausible theory of justification must satisfy.[30] The position that justification has a historical dimension amounts to the view that beliefs are not justified merely by facts about the moment of belief formation or evaluation. Rather, beliefs can be (at least in part) justified by historical facts, i.e., facts about what obtained prior to the moment of belief formation and evaluation. The justificatory status of a subject's belief that *p* at time t can, to some extent, be

[30] Unless otherwise indicated, "justification" refers to discuss "doxastic justification".

determined by facts about what obtained prior to *t*.[31] Historical accounts of justification stand in contrast to ahistorical or time-slice views of justification. These views claim that justification solely depends on facts about the moment of belief evaluation. In other words, whether a subject's belief that *p* is justified at time t is wholly determined by facts about time *t*. Time-slice theorists argue that we can determine whether a subject's belief is justified by taking a "snapshot" of the moment of belief evaluation; proponents of historical accounts of justification argue that we need to look at what obtained prior to that time.

At the outset, it should be noted that the claim that any plausible theory of justification is historical is not deeply divisive or very controversial. The historical versus time-slice debate neither characterizes, nor tracks onto, the divide between internalists and externalists. Neither does it explain or track the divide between foundationalists and coherence theorists. As the following discussions will show, reliabilists and evidentialists alike can, with good reason, endorse the claim that justification has an important historical dimension. [32] To demonstrate this, I will rehearse the following argument in favor of historical accounts of justification in terms of reasons, not processes.[33]

Contrasting doxastic and propositional varieties of epistemic justification reveals the necessary historical dimension of the former variety.[34] As I mentioned previously, my claim is that any plausible

---

[31] I will proceed with this minimal, conservative account of what historical facts are: historical facts are facts concerning what occurred prior to the moment of belief formation and evaluation. Epistemologists could, of course, disagree about whether the historical facts that are relevant to justification are solely facts about a subject's mental states, or whether historical facts can include facts about what obtained outside a subject's psychology. This would amount to a debate between internalists and externalists. However, neither conception of historical facts is inconsistent with a historical theory of justification. For an interesting discussion of different ways of conceptualizing historical justificational facts, see Thomas Kelly 2016 (44-51).

[32] Time-slice accounts of justification are often associated with internalism. For example, prominent defenders of time-slice views include Roderick Chisholm (1989, 59-60), Richard Feldman and Earl Conee (2004, 55, 101) – all defenders of internalism. However, see Jeremy Fantl (2019) for discussions of how one could craft a time-slice externalist position (780-2), and for a defense of evidentialism as a historical account of justification (784-787).

[33] An overview of the historical commitment cashed out in terms of processes can be found in section 4.1.

[34] There are, of, course, other arguments in favor of historical theories. Goldman's argument from "preservative memory" is discussed in section 5, footnote 33 (2009, 323). In short, his argument is that historical accounts are necessary given we are sometimes justified in believing that *p* even when we don't remember the evidence that originally caused and justified our belief that *p*.

account of *doxastic* justification must be historical. When a subject has propositional justification, they are *in a position* to form a doxastically justified belief that p, perhaps because they have sufficiently good reasons to believe that p. A subject with propositional justification has a *justifiable* belief. In contrast, a subject is doxastically justified when they properly base their belief that p on sufficiently good reasons. Consider two subjects, Kate and Stephen.[35] Both believe the proposition q, and both believe and have doxastic justification for the propositions *if p, then q*, and *p*. Kate believes that q on the basis of her beliefs that *p* and *if p, then q*. Wishful thinking prompts Stephen to believe that *q*. The basis for Stephen's belief is epistemically defective despite the supporting evidence he has. Kate is doxastically justified in her belief that q while Stephen merely has propositional justification for his. This is because Kate's belief that *q* is properly based on her evidence while Stephen's is not. What does this case demonstrate? It demonstrates that proper basing is a necessary condition of doxastic justification.[36, 37]

Once we have accepted that proper basing is a necessary condition of doxastic justification, we need only reflect on the causal nature of the basing relation to see that this condition entails that justification has an important historical dimension. When we say that Kate properly bases her belief that *q* on her beliefs that *p* and *if p, then q*, we are saying that Kate's belief that *q* depends on those supporting beliefs. A belief is dependent on another only if the two are causally related. Kate's belief that q depends on her beliefs that *p* and *if p, then q* insofar as the former two beliefs caused her to form the belief that *q*, and/or cause her to continue believing that *q*.[38]

---

[35] This argument is made by Hilary Kornblith (1980, 602). While he does not construct his argument using the propositional vs. doxastic justification distinction, the introduction of these terms does not change the nature of the argument. Harman (1973, 30-33) makes a similar argument.

[36] Importantly, *proper* basing, not basing *simpliciter*, it a necessary condition for doxastic justification. For a discussion of why this qualification is necessary, see Paul Silva (2015).

[37] As mentioned previously, this argument was run in terms of evidence and reasons (rather than processes) to demonstrate that this historical commitment can cut across different theories of justification. That said, further discussion of the historical commitment using the language of belief-forming processes can be found in section 4.1.

[38] As Kornblith notes, recognizing the causal nature of belief dependence allows us to appreciate "an important insight of foundationalism" (1980, 602). Some argue, for example Korcz 2000, that not all instances of proper epistemic basing involve causation. The thought is that a subject could have already formed a belief that *p*, and then realize that she has good reason *r* to believe *p*. In this case, she forms a meta-belief to that effect. Korcz argues that this meta-belief is a justification-conferring basis for the subject's belief that *p* even though the meta-belief did not cause the subject's belief

Given that (i) proper basing is a necessary condition of doxastic justification, (ii) basing is a causal relation, and (iii) causation is a temporally extended process, doxastic justification does not merely supervene on facts about the moment of belief evaluation. (i) is the widely accepted way of interpreting the distinction between propositional and doxastic justification. (ii) is the widely accepted, "standard" interpretation of the epistemic basing relation (Korcz 2000, 526).[39] Insofar as we accept that causes temporally precede their effects, we must also accept (iii).

Before concluding this section, I want to note that the foregoing argument makes perspicuous an explanatory desideratum that any plausible theory of doxastic justification must satisfy. Ultimately, what the discussion of proper basing demonstrates is that investigations into a belief's justificatory status are investigations into, broadly speaking, what we might term (non-)accidentality explanations. The rough idea is the following: If the subject's true belief is improperly based on their evidence, like the case of Stephen above, then the subject's belief is merely accidentally true, and consequently, unjustified. If a subject's true belief is properly based on good evidence, like the case of Kate above, then the subject's belief is non-accidentally true, and consequently, justified. In the former case, the truth of the belief is a matter of luck; in the latter case, the truth of the belief is not a mere accident. These determinations of luck and non-accidentality are fundamental to evaluations of justification and must be accounted for by any acceptable theory of doxastic justification. As such, epistemologists ought to commit to the following desideratum for theories of justification:

> **The Explanatory Desideratum for Theories of Doxastic Justification:** The correct theory of doxastic justification will accurately account for all parts of the (non-)accidentality explanation for the truth or falsity of a subject's belief that *p*. In other words, all parts of the (non-)accidentality explanation for the truth or falsity of a subject's belief that *p* must properly feature in the correct theory's account of the relevant belief's justificatory status.

---

that *p*. However, this meta-belief is clearly part of the explanation of what causes the subject to continue believing that *p*. Given the meta-belief is part of what casually sustains the subject's belief that *p,* we need not concede that the epistemic basing relation is not a causal relation in such cases.

[39] See also Fantl (2019) for a discussion of the basing relation as a "diachronic requirement" (784-787).

The argument above demonstrates that satisfying the explanatory desideratum requires accepting that justification has an important historical dimension. Facts about the moment prior to belief formation or evaluation feature in (non-)accidentality explanations, and therefore must be part of the justificatory story – i.e., these historical facts must be part of the explanation a theory of doxastic justification gives for why a belief is (un)justified.

With this in mind, I will transition to my discussion of Mercier and Sperber's interactionist theory of reasoning. I conceive of their theory as new, epistemically relevant data about what explains why subjects have the beliefs that they do. In particular, I argue that it is data about why subjects form non-accidentally true beliefs in social, deliberative settings, and as such, it must be part of the justification story in relevant cases. In light of the explanatory desideratum, I explore how theories of doxastic justification can accommodate the epistemic consequences of Mercer and Sperber's view.

3. Mercier and Sperber's Interactionist Theory of Reasoning

In *The Enigma of Reason*, Mercier and Sperber (2017) reject the "dogma" they argue lies at the heart of traditional psychological and philosophical theories of conscious reflective reasoning, and develop and defend a novel "interactionist theory" (9, 21, 182-183).[40] Some clarification of what is meant by "theories of reasoning" is undoubtedly necessary. First, by reasoning, Mercier and Sperber mean exercising the ability to produce, consciously entertain, and evaluate reasons for and against believing particular propositions and performing particular actions.[41] Second, on Mercier and

[40] This book is a continuation and development of the view Mercier and Sperber present in "Why do humans reason? Arguments or an Argumentative Theory" (2011) and discuss in "Reasoning as a Social Competence" (2012).

[41] Some might be concerned that this notion of reasoning is oddly restrictive and argue that we surely engage in unconscious reasoning. I would argue this concern is more terminological than substantive. Mercier and Sperber believe that reasoning necessarily involves the use of psychological reasons, i.e., representations that represent information as a reason to support a particular conclusion (2017, 119-120). While they think there is undoubtedly unconscious information processing that results in the formation and revision of beliefs, they do not think that the representations involved in such processes count as psychological reasons. They think it more likely that unconscious belief-revision processes just exploit regularities between different pieces of represented information. That said, my argument does not turn on this debate. So long as the reader grants that that there are psychological processes like reasoning in Mercier and Sperber's sense, and that these processes can be properly analyzed using the relevant parts of their interactionist theory that I outline below, then my argument faces no objection on this score.

Sperber's view, a theory of reasoning is (at least) a two-part explanation of that phenomenon. A theory of reasoning explains (i) what reasoning is for and (ii) how it works. The traditional dogma that is Mercier and Sperber's primary target solely stakes out a position concerning (i): what reasoning is for. According to this traditional view, the capacity to produce and consciously entertain reasons for and against believing particular propositions or performing particular actions is a capacity that helps us, as individuals on our own, form more true beliefs and act more pragmatically. Mercier and Sperber term this orthodox position the "intellectualist theory" (Mercier and Sperber 2017, 330-331).[42]

Why reject the intellectualist theory, the traditional "dogma" that, as Mercier and Sperber put it, assumes that "the job of reasoning is to help *individuals* achieve greater knowledge and make better decisions" (Mercier and Sperber 2017, 4; emphasis mine)? Mercier and Sperber argue that empirical research into the strengths and weaknesses of this cognitive capacity reveal reasoning as primarily the exercise of a "social competence" (11). They are not claiming that reasoning fails to bring about "intellectual benefit" for individuals (11). Rather, their position is that reasoning brings about these individual epistemic and pragmatic benefits through "interactions with others" (11). Reasons, they argue, are produced for "social consumption": for convincing others and for justifying ourselves to others (127). When reasons are produced for this purpose, exchanged and evaluated in collaborative dialogic contexts, interlocutors more reliably reach true, well-reasoned conclusions. The epistemic gains of reasoning are explained by the interaction between interlocutors in dialogue, not interlocutors' individually reliable reasoning faculties. In other words, reasoning works well when collaboratively employed in the capacity's normal conditions: social and dialogical conditions (247). Let us turn our

[42] The intellectualist theory admits of diversity. Two philosophers or psychologists could agree on this position regarding (i) but disagree about (ii). For example, one could think that reasoning is an individually exercised and individually beneficial capacity while arguing that reasoning functions using mental models (for example, Johnson-Laird and Byrne 1991). Distinctly, one could think that reasoning is an individually exercised and individually beneficial capacity while claiming that reasoning works using logical rules of inference (for example, Rips 1994). Therefore, the individualistic orthodox conception of reasoning, the target of Mercier and Sperber's criticism, is more accurately thought of as a cluster of views that all characterize reasoning as an individually exercised cognitive capacity that practically and epistemically aids individuals through "solitary ratiocination", rather than a single unified view (Mercier and Sperber 2017, 218).

attention now to the empirical research that Mercier and Sperber argue supports this recasting of reason as a social competence.

Since the mid-1970s, empirical research into systematic errors in human reasoning has resulted in disagreement about the extent of human rationality – a debate sometimes termed the "rationality wars" (Mercier and Sperber 2017, 21). The relevant studies have demonstrated that human reasoners systematically make certain kinds of errors. Strikingly, reasoners commonly fail to abide by certain rules of probability and logic. For example, Amos Tversky and Daniel Kahneman (1983) conducted research showing that human reasoners predictably fall prey to the conjunction fallacy (i.e., fallaciously believing that there is a greater probability of "A and B" than there is of just "A" or "B"). Moreover, research suggests that human reasoners fail at alarming rates to abide by logical rules governing conditionals (Mercier and Sperber 2017, 28; Byrne 1989; Evans 1989). Studies also show that human subjects are often mistaken about their own reasons for belief and action, and frequently confabulate such reasons (Wason and Evans 1975; Nisbett and Wilson 2002; Lucas and Ball 2005; Halberstadt and Wilson 2008; Hauser *et al* 2007; Carruthers 2011). Additionally, reasoners fall prey to the confirmation, or as Mercier and Sperber prefer "myside", bias: the tendency to single-mindedly pursue and consider only that evidence confirming the belief the reasoner already holds (Mercier and Sperber 2017, 213; Johnson-Laird and Byrne 2002; Stanovich and West 2008; Stanovich, West, and Toplak 2013). In addition to this bias in evidence and reason identification, individual isolated reasoners tend to be limited in their ability to produce compelling reasons for their own beliefs and to anticipate counterarguments (Mercier and Sperber 2017, 223; Kuhn 1991; Nisbett and Ross 1980; Perkins 1985). These last two traits of reasoning can lead to problematic "epistemic distortions": "overconfidence, polarization, belief perseverance" (Mercier and Sperber 2017, 246). On one side of the debate are thinkers who take this evidence of widespread systematic errors in human reasoning as cause to be pessimistic about human rationality (e.g., Piattelli-Palmarini 1994; Kahneman and Tversky 1973); on

the other, there are those who think there are alternative, more optimistic explanations (e.g., Pinker 1997; Gigerenzer 1991 and 1998).

Mercier and Sperber agree these empirical findings are relevant to investigating the purpose and mechanics of reasoning. We learn, after all, about how perception works by studying errors in perception, i.e., illusions (Mercier and Sperber 2017, 22). Similarly, errors in reasoning should help us learn about how reasoning works. If one is a proponent of the intellectualist theory of reasoning, then one believes that reasoning is for isolated individual use. It ought to work well when used in isolation by individuals. It is understandable, therefore, why the experimental results discussed above might prompt a proponent of the intellectualist theory towards pessimism about human rationality. Mercier and Sperber, like others sympathetic to the evolutionary psychologist's research program, reject this rationality-pessimistic explanation of the data. Their reasons for doing so are straightforward: a rationality-pessimistic explanation of the data involves claiming that we have an ill-adapted cognitive capacity, which, from an evolutionary perspective, they think is hard to accept:

> [On the framework of the traditional intellectualist theory of reasoning,] failures of reasoning are lazily explained by various interfering factors and by weaknesses of reasoning itself. Again, this doesn't make much evolutionary sense. A genuine adaptation is adaptive; a genuine function functions. (Mercier and Sperber 2017, 331)

If we accept the insights of evolutionary psychology, Mercier and Sperber argue, we should think that an ill-adapted competence is puzzling and likely misdescribed. How do we go about properly describing the competence? A well-adapted competence, they argue, is not an ability to perform a particular function under *any* conditions. Rather, a well-adapted competence functions well under *normal* conditions.[43] Mercier and Sperber propose we must determine the conditions under which

[43] To clarify, Mercier and Sperber use the following example: we do not think that our lungs are ill-adapted for their functional role because they fail to operate under water. Under water environments are abnormal conditions for human lungs. Our lungs are adapted to function well in *normal* conditions: "the earth's atmosphere at ground level" (Mercier and Sperber 2017, 247). See Ruth Millikan (1987, 34) for more discussion of biological mechanisms having normal operating conditions.

reasoning works well or optimally. *Then*, they claim, we can determine (i) what reasoning is for and (ii) how it works. One way of clarifying Mercier and Sperber's claim is as follows: the mistake of the traditional intellectualist approach is that it reverses the proper order of analysis. One does not start by stipulating that a certain system has a particular function, then look at evidence of that system failing to perform that function well as cause for thinking that it is a poorly operating system. Rather, one looks at evidence of when and under what conditions the system works well, then one determines what function the system was adapted to perform.

Under what conditions does reasoning work well? Mercier and Sperber argue that the data demonstrates reasoning works well under social, cooperative dialogical conditions, i.e., when reasoners engage with, justify themselves to, and disagree with other reasoners in an effort to coordinate beliefs and actions (Mercier and Sperber 2011, 247; 2017, 227, 183-186). As noted above, individual isolated reasoners are not particularly adept at producing compelling reasons for their beliefs or anticipating counterarguments. Mercier and Sperber argue that the experiments at the heart of the rationality wars are not ones in which participants are in a "typical dialogical context" (Mercier and Sperber 2017, 227). In contrast, experiments that place participants in dialogical contexts demonstrate that reasoners are quite adept at producing reasons to justify themselves *to their interlocutors* and crafting counterarguments to the positions held by their interlocutors (Mercier and Sperber 2017, 228; Resnick et al. 1993, 362-363; Kuhn, Shaw, and Felton 1997). Therefore, by the lights of Mercier and Sperber's interactionist approach:

> "…the normal conditions for the use of reasoning are social, and more specifically dialogical. Outside of this environment, there is no guarantee that reasoning acts for the benefits of the reasoner. It might lead to epistemic distortions and poor decisions. This does not mean that reasoning is broken, simply that it was taken out of its normal conditions." (2017, 247)

When deployed in dialogic social contexts, Mercier and Sperber argue the characteristics that some have taken to be evidence of reasoning's flaws – such as the ability to primarily produce reasons in

defense of one's beliefs and the inability to anticipate counterarguments – are repurposed into collaborative strengths. The confirmation or "myside bias" is less likely to result in "epistemic distortions" insofar as it functions as an effective division of cognitive labor (Mercier and Sperber 2017, 221). The bias divvies up the group's investigative work. Finding arguments and reasons in defense of a position with which one does not agree is cognitively costly. If interlocutors focus on producing arguments for their own position, the deliberative group can then, in a sense, pool their cognitive resources and increase their chances at arriving at a justified conclusion given they are able to cover more evidential ground. Moreover, the biases that motivate the isolated individual are more likely to be corrected in dialogic contexts insofar as they are kept in check by the countervailing biases of interlocutors. Deployed in a collaborative dialogical context, the confirmation bias doesn't leave the individual reasoner only with arguments in favor of the position she came into the discussion believing. Rather, she has her own arguments, objections to those arguments leveled by her interlocutor, arguments for contrary positions, and her own evaluation of those contrary arguments.

With this understanding of the conditions of proper functioning in place, Mercier and Sperber's interactionist theory takes shape. In an intermediary sense, the function of our individual reasoning processes is to convince others, to justify ourselves to others, and to evaluate others' arguments. When this intermediary function is carried out in sufficiently collaborative dialogic contexts, it achieves its ultimate function: our epistemic benefit.[44] Reasoning is not a capacity for individual epistemic gain through "solitary ratiocination" (Mercier and Sperber 2017, 218).[45] When

[44] The language of intermediary and ultimate function is not Mercier and Sperber's, but rather my own. This decision is meant to highlight those features of their view that are of particular interest to epistemologists.

[45] Given this paper concerns epistemic justification, some may be concerned that an implausible consequence of the interactionist theory is that we are never justified in beliefs that we form through solitary, conscious reflective reasoning. Insofar as Mercier and Sperber are arguing that reasoning is the exercise of a social competence, and therefore unreliable when exercised in isolation, it seems like they are potentially committed to saying that solitary ratiocination always yields unjustified beliefs. But surely, one might argue, we have all sorts of justified beliefs that are the output of solitary ratiocination, and therefore Mercier and Sperber's theory is implausible. To this objection I offer two responses: one that is Mercier and Sperber's, one that is my own. First, Mercier and Sperber note that the problem isn't solitary ratiocination per se, but solitary conscious reflective "reasoning that *remains* solitary" (2017, pp, 249). Second, we can accept that

deployed collaboratively in social dialogic contexts, this socially directed competence brings about epistemic goods. The deployment of our cognitive capacity to produce justifying reasons and evaluate counterarguments is prompted, or tripped, when we are confronted with disagreement or doubt from others. It is the need to convince others—to convince them that what we say is true, or that we are justified in acting a certain way—that triggers our reasons-producing and evaluating faculties.[46] When we combine these cognitive efforts, we jointly pave our way to epistemic goods.[47]

Central to Mercier and Sperber's defense of their interactionist theory is the contention that a proper theory explains the capacity's characteristics as integral to the competence and success of the system, not as errors in functioning. As previously mentioned, on the interactionist theory, confirmation biases are not epistemic disadvantages insofar as countervailing confirmation biases balance out in interactive dialogic contexts, the normal conditions of reasoning. In a similar vein, solitary reasoners' limited ability to produce reasons and evaluate their own arguments is not a fault

---

reasoning is the exercise of a social competence, and therefore reliable when deployed in social contexts, while also maintaining that it can be cautiously used to form justified beliefs in isolation. An analogy will help illustrate why this is the case. Our visual perceptual faculties have evolved to form reliable beliefs about our environment in settings with adequate lighting. As such, the human visual perceptual system is reliable in contexts with adequate light. That said, we are often justified in forming beliefs using this competence even in settings with dim lighting. We ought to exercise greater epistemic caution when doing so, and the resulting beliefs likely have a lower justificatory status than those that are the outputs of the same competence deployed in the proper context of use. All to say: claiming that an epistemic competence has evolved to functional reliably in a particular environment does not entail that we never form justified beliefs when we use the competence outside that particular environment.

[46] At this juncture, one might have several concerns. It is, of course, outside the scope of this paper to rehearse all of Mercier and Sperber's arguments and defenses of their interactionist theory. That said, a few clarifications will help to address some immediate objections. First, one might be concerned that this theory presupposes too much disagreement. Is disagreement really this pervasive? To this worry, it is important to highlight that in a dialogical context, one might be called on to justify oneself to one's interlocutor even if one's interlocutor doesn't necessarily disagree (for example, they may not have firm beliefs about the issue under discussion). Furthermore, Mercier and Sperber note that the anticipation of disagreement can be sufficient to trip our reason-producing faculties (2017, 248). Finally, I want to guard against an overly bombastic, polarized notion of the disagreement and clash that Mercier and Sperber take to be essential to the normal condition in which reasoning functions well. With a sufficiently expansive notion disagreements, clashes of idea, and interactive dialogue, we can see that the proposed normal conditions of interactionist reasoning are pervasive. They obtain when talking about quotidian topics – when deciding with friends where we should go for lunch or discussing the best route to take the airport – and conversations that have a clear argumentative flavor – when disagreeing about new policy or debating how to interpret research findings.

[47] As mentioned previously, theories of reasoning have two primary components: (i) an explanation of the function of reasoning, (ii) and explanation of how reasoning works. Regarding (ii): in *The Enigma of Reason*, Mercier and Sperber go on to develop a modular account of reasoning (2017, 129-174). I will not explicate the details of this part of the view largely for concision's sake, but moreover because no part of my argument turns on accepting their claim that reasoning is modular.

of reasoning, rather it is a cost-saving part of the competence. Dialogic contexts facilitate the quick exchange of reasons and arguments. One need not be particularly skilled at anticipating counterexamples if they are provided by one's interlocutor, but one needs to be capable of producing arguments for one's belief when responding to direct questions raised by one's deliberative colleagues. Given that the identification and evaluation of reasons and arguments is cognitively costly, Mercier and Sperber argue it is fitting that such capacities are reflexively exercised only when the need arises. Similarly, biased reasoning is cost-effective when it is a means by which participants in a dialogical context divide up fact-finding, i.e. argumentative, tasks. As such, the negative epistemic consequences associated with reasoners' individual poor abilities to anticipate counterexamples, formulate extensive arguments, and consider evidence in an unbiased manner are put to good epistemic use in the consistent back-and-forth of social deliberation.

4. A New Kind of Epistemic Reliance: Extended Interactive Justification-Conferring Processes

Let's say we accept the positions defended in the previous two sections: (i) the view that any plausible theory of justification must be historical in light of the explanatory desideratum for theories doxastic justification, and (ii) Mercier and Sperber's interactionist theory of reasoning. How ought we account, as epistemologists, for the epistemic gains that come from dialogical deliberation? Dialogical deliberation can surely be part of the history of a belief, but ought we conceive of it as a part of the belief's epistemically relevant history? In other words, ought we think of dialogical deliberation as part of the (non-)accidentality explanation for the truth or falsity of some beliefs? Can dialogical deliberation be an element of a belief's history that impacts justification? In this section, I argue that we should answer all these questions in the affirmative. I start by showing how processes reliabilism can and should straightforwardly accommodate the interactionist theory. Process reliabilism can satisfy the explanatory desideratum if Mercier and Sperber's interactionist account of reasoning is true by conceiving of (some) justification-conferring processes as extended and interactive. In the next

section, I will argue that other theories, in particular internalist theories, are unable to satisfy the explanatory desideratum in light of Mercier and Sperber's theory.[48]

4.1 In Defense of Extended Interactive Belief-Forming Processes

Process reliabilism, the view that justified beliefs are the outputs of reliable belief-forming processes, can elegantly account for epistemically relevant events that occur prior to the moment of belief formation. This is in large part because the theory was originally explicitly formulated to account for the intuition that a belief's justificatory status is impacted by that which "*causally initiates*" or "*sustains*" it (Goldman 1979, 8).[49] The general strategy of process reliabilism's inaugural text – Alvin Goldman's "What is Justified Belief?" – is to object to time-slice theories of justification by arguing that they yield unintuitive verdicts in cases where a belief is genetically corrupt. The case of Stephen from section 2 is just such a case. Recall, Stephen has a propositionally justified belief that *q* insofar as he justifiedly believes both that *p* and *if p, then q*. He is not doxastically justified in his belief that *q* because his belief that *q* is not based on these former two beliefs, but rather is the product of wishful thinking. Forming a belief through a process of wishful thinking is unreliable, and therefore Stephen's belief is doxastically unjustified. If Stephen's belief is not formed by a process that reliably aims at the truth, it can only be accidentally true; if a belief is only accidentally true, it cannot be justified. This is the basic argument for accepting process reliabilism.

Before moving on, I just want to highlight that Goldman's process reliabilism, and the articulation of the view accepted by most of its proponents, maintains that the reliability of the *entire* process of belief formation is relevant to a belief's justificatory status. Consider this variant on

[48] The interactionist theory is, of course, a relatively new view and as such it would be contentious to claim that the theory that most elegantly incorporates its insights is preferable. However, should empirical research and analysis of that research continue to affirm the central tenets of Mercier and Sperber's position, theories of justification should be able to account for the way in which reasoning is the exercise of a social competence.

[49] Goldman termed this early iteration of the view "*Historical Reliabilism*" (1979, 14). More precisely, the view is that a belief is justified if and only if it is the output of a reliable process, or a *conditionally* reliable process. A process is conditionally reliable when its inputs are beliefs. In such cases, the reliability of the process is conditional on the truth of the process inputs (1979, 13).

Stephen's case. As before, Stephen has a belief that *q* and justifiedly believes that *p* and *if p, then q*. As before, Stephen doesn't base his belief that *q* on the two latter beliefs. Rather, he bases his belief that *q* on the beliefs that *r* and *if r, then q*. Stephen reasons perfectly from these premise-beliefs to the belief that *q* – flawlessly executing modus ponens. However, Stephen's beliefs that *r* and *if r, then q* are the products of unreliable, wishful thinking. Despite his perfect execution of modus ponens reasoning, Stephen's belief that *q* is unjustified because the premise-beliefs were unreliably formed. Whether or not his belief that *q* was formed reliably depends on whether the premise-beliefs were themselves reliably formed.[50] This insight prompts Goldman to reject what he terms "*Terminal-Phase Reliabilism*", the view that justification is conferred only by the reliability of the process that occurs that the very end of belief formation (16). Justification, according to the reliabilist, requires that the "*entire* history of the [belief-forming] process be sound (i.e., reliable or conditionally reliable)" (Goldman 1979, 16).[51]

If Mercier and Sperber's interactionist theory is correct, we should accept the following: Just as the processes by which Stephen formed his premise-beliefs are relevant to the justificatory status of his conclusion-belief, interactive deliberation is relevant to the justificatory status of beliefs formed in dialogical contexts. When a subject engages in dialogical deliberation and forms a belief as a result of that engagement, part of what explains whether or not the belief was reliably formed is the subject's interaction with their interlocutors.[52] According to the interactionist theory, in these kinds of cases part of what accounts for the (un)reliability of subjects' reasoning is the (lack of) collaborative

[50] Just to note: this insight can be formulated in reliabilist or evidentialist terms. The reliabilist can say, "whether Stephen's conclusion-belief is justified is not just a function of the reliability of his reasoning from premises to conclusion, but the reliability of the processes by which he came to believe those premises as well". The evidentialist can say, "whether Stephen's conclusion-belief is justified is not just a function of whether Stephen properly bases his belief on the evidence, where evidence is construed as the premises of his argument. The justificatory status of his conclusion-belief also depends on whether the premises of his argument are sufficiently evidentially supported."

[51] Two clarifications are necessary. First, strictly speaking, Goldman asserts that "Terminal-Phase Reliabilism" is a "brand of justifiedness", essentially a variety of justification that is "not so closely related to knowing" (1979, 16). If one should posit that there are different varieties of justification, let me stipulate here that my focus is on the variety that I view as related to knowing. Second, *conditionally* reliable processes are belief-forming processes that take beliefs as their inputs (Goldman 1979, 13).

[52] Like Goldman, I posit that justification is not merely conferred by processes that form beliefs but processes that sustain them. However, for clarity's sake, I will continue using only the language of belief formation.

engagement with interlocutors. As such, any historical process reliabilist should accept that the dialogical deliberation is part of the epistemically relevant history of the belief, and therefore is one of the processes, or part of the process, that confers justification. Once accepted, the process reliabilist must also grant that justification-conferring processes do not solely occur within individual subjects' cognitive systems. Rather, justification-conferring processes can extend beyond this boundary to include the cognitive systems of deliberative interlocutors. In cases of beliefs produced or sustained by dialogical deliberation, the epistemically relevant, justification-conferring process does not only occur in the cognition of the individual subject whose belief is under evaluation. Rather, the justification-conferring process extends beyond the individual's cognition, occurs in part in the cognitive systems of the deliberative dialogue's participants, and as such encompasses the back-and-forth of discussion. I will term this kind of process an *extended interactive belief-forming process.*

4.2 An Illustrative Example

Let's take a case where there is sufficient engagement and adequate collaboration such that the deliberative dialogue is reliable, and (most of the) participants in the discussion come away with a true, justified belief. Imagine a hiring committee composed of six members charged with assessing a pool of applicants and identifying the most qualified, best-suited candidate. After each of the six committee members reviews the applicants' dossiers, the committee convenes to deliberate and make a final recommendation to the department. Some members arrive at the meeting with an unjustified belief regarding the most qualified candidate; some don't yet have a settled view. In the back-and-forth of discussion, all are asked to defend the claims they make. The subsequent evaluation of reasons and arguments prompts the production of stronger arguments, and the dismissal of certain applicants as contenders. For example, we can imagine that the following exchange is characteristic of the committee's deliberation.

Madeleine: Michelle went to a top-ranked MBA program.

> *Kevin*: That's true, but Michelle has minimal work experience. Jesse went to a similarly ranked program and has been working in a company like ours for five years.
>
> *Tanja*: Wouldn't it be better to hire someone we can train from the ground-up?
>
> *Madeleine*: That's a reasonable point. But Jesse has worked and been successful at a variety of companies. I think that demonstrates an ability to adapt well to new settings.
>
> *Tanja*: In that case I think we should be concerned that Jesse won't stick around very long if hired. It does seem like they've moved through employers quickly, and we've made the mistake before of hiring people who don't stay long enough.
>
> *Kevin*: Jesse indicated during our initial phone interview that they're relocating to the area to be close to family. I don't think we have to be concerned about that.
>
> *Madeleine*: Also, Jesse has worked at previous employers for no less than three years, which is comparable to our employee turn-over rate for this kind of position.

The committee's deliberation continues in this manner. Most of the committee's members come to be justifiably convinced that Jesse is the most qualified, best-suited candidate.

What is epistemically relevant to the justificatory status of the committee members' beliefs that Jesse is the most qualified candidate? What explains why the committee members' belief-forming processes are reliable? If we accept the interactionist theory, for any individual committee member it is not merely their own isolated cognitive activity that determines the reliability of the process by which their belief was formed. Rather, their interlocutors' cognitive activity, and the interaction between that activity and their own individual processing, is also relevant. Tanja's critical evaluation of considerations offered by Kevin prompts Madeleine to counter with a defense of Jesse's candidacy. This in turn prompts Tanja to produce further reasons to be concerned with Jesse's application. The interactionist theory of reasoning tells us that these are not arguments, reasons, and evaluations that the committee members would have necessarily produced working in isolation. Nor is it the case that the short processes yielding these intermediary arguments, reasons, and evaluations are individually reliable when isolated from the interaction of the dialogue. Rather, these cognitive feats are tripped by one another's engagement and contributions; and moreover, these individual pieces of reasoning

contribute to reliability when triggered by, and deployed in, this collaborative context.[53] Therefore, the relevant justification-conferring process is social. For beliefs formed in this context, the processes extend beyond any one individual's cognition and include other interlocutors' cognitive activity as well as interactions between those cognitive events.

4.3 Problems for Alternative Individualistic Approaches

Consider what happens if we try and craft a more individualistic epistemic accounting of the interactionist theory of reasoning. A more individualistic analysis would isolate and evaluate solitary acts of reasoning that occur during the discussion. Take one of Tanja's objections to Jesse's candidacy: the idea that the committee ought not hire Jesse because they won't stay at the company for a sufficiently long period of time. On the framework of the interactionist theory, the individual, isolated process of ratiocination that produced this small argument is likely not reliable when considered in isolation. Perhaps Tanja came to the discussion with an unjustified false belief about who the best candidate is, and this individual, isolated process of ratiocination is steered by the confirmation bias. As such, it is unreliable when evaluated on its own. Similar descriptions are true of other participants' segmented, individual processes. How then do we then explain that the committee members end up with justified beliefs at the end of the deliberation? Perhaps an individualist analysis would suggest that we do so by a process of aggregation: whether or not a committee member's belief is justified depends on the reliability of the aggregate of their individual reasoning processes during discussion.

[53] One might argue that attaching justificatory value to conscious explicit processes of argument production and evaluation runs counter to reliabilism's deep externalists commitments. However, a committed externalist process reliabilist can (and should) accept that a socially interactive justification-conferring process can yield a justified belief at the end of group deliberation, even if interlocutors cannot recite the reasons discussed by the group in favor of the belief formed. The justification is conferred by the reliability of the dialogue, not by the reasons possessed by the interlocutors of the end of the discussion.

This is an individualist approach insofar as it limits the belief-conferring processes to the cognition of the subject whose belief is under evaluation.

However, it is not clear why merely aggregating individually unreliable processes would yield a sufficiently reliable justification-conferring process. Moreover, even if individual pieces of ratiocination aggregated together to yield a sufficiently reliably process, this proposal should still be rejected. Such an account would characterize committee members who form justified beliefs as having done so *in spite of* the individual pieces of ratiocination that are unreliable in isolation. Tanja, for example, would be justified *in spite of* engaging in unreliable biased reasoning that produced an intermediate argument against Jesse's candidacy. This analysis fails to satisfy the explanatory desideratum outlined previously. Why? Because on Mercier and Sperber's view, part of what explains Tanja and other committee members' success is that she engaged in this biased process of reasoning. Interlocutors that engage in group problem solving are epistemically successful *because of* these individual pieces of ratiocination that are unreliable in isolation. Although unreliable in isolation, they are, as discussed in section 3, integral for reliability in collaborative dialogue.[54] As such, the individualistic analysis under discussion misrepresents epistemically relevant events in its accounting of the non-accidentality explanation.

Some may argue that an individualistic explanation of the case is still available. One could propose the following: To see what confers justification on any individual committee member's belief, look at the precise moment where the committee member forms the belief that Jesse is the most qualified candidate. Although the back-and-forth of dialogue is interesting, justification is conferred

[54] I want to emphasize that this "in spite of" vs. "because of" distinction is not a mere shallow point about language. To say that a committee member is justified in spite of these individually unreliable segments of reasoning is to be committed to saying that the subject in question would be epistemically better off, i.e., would have a more justified belief, had they not occurred. Mercier and Sperber's theory undermines this precise point. More abstractly, "in spite of" and "because of" explanations are substantive and appropriate within debates about epistemic justification, as justification is matter of applying normative epistemic evaluations to what explains why a subject believes as she does.

by the reliability of this end-of-discussion belief-forming process. Like the strategy above, this is an individualist analysis insofar as this process occurs wholly within the cognition of the individual whose belief is under evaluation. However, this solution is not plausible as it fails to fully accommodate the historical dimension of justification defended in section 2. This individualist strategy doesn't account for cognitive activity on which the belief in question epistemically depends. According to the interactionist theory, the individual segments of solitary ratiocination (that are unreliable in isolation) are integral to the epistemic success of the collaborative dialogue, and as such, they are part of the epistemically relevant history of belief formation. In totally removing these cognitive events from the explanation, this individualist approach is embracing a kind of terminal-phase reliabilism that, as discussed above, violates the reliabilist's commitment to a historical theory of justification and the explanatory desideratum.

The committee members' beliefs that Jesse is the most qualified are non-accidentally true because of one another's interaction and engagement, and therefore, that interaction and engagement is part of the epistemically relevant history of belief formation. That which is epistemically relevant to a belief's history bears on its justificatory status. A process reliabilist should account for this by positing that, in cases of dialogical deliberation, the relevant justification-conferring process extends beyond the individual to include the implicated cognitive activity of their interlocutors. In such cases, the relevant justification-conferring process is interpersonally, interactively extended.

4.4 Extended Interactive Justification-Conferring Processes as Truly Social Epistemology

I acknowledge that this is a significant departure from the way in which reliabilists have traditionally understood the extent of justification-conferring processes.[55] However, my project is

[55] Indeed, when first engaging with the question of how reliabilists should conceive of the "extent" of belief-forming processes, Goldman, "with some hesitation", suggests that we "restrict the extent of belief-forming processes to '*cognitive*' events, i.e., events within the organism's nervous system" (1979, 13). His primary reason for doing so is that justification is an evaluation of "how a cognizer deals with his environmental inputs" and we best capture this insight by restricting the relevant processes to individual subjects' cognitive systems. I will respond to this concern in the next section.

following social epistemology's tradition of acknowledging the deeply social aspects of our epistemic lives and theorizing about how best to accommodate these insights into our normative epistemic framework. Consider the parallels between my argument and those Sanford Goldberg makes in his book, *Relying on Others*. Goldberg's central aim in this text is to defend his "extendedness hypothesis": the view that the relevant justification-conferring process for beliefs formed on the basis of testimony includes not only the hearer's own cognitive processes, but "the cognitive processes implicated in the production of that testimony" (Goldberg 2010, 79). On Goldberg's view, whether or not the belief I form on the basis of testimony is justified depends on more than just the reliability of my own belief-forming cognitive processes. It also depends on the reliability of the relevant cognitive processes occurring in the testifier. Goldberg's argument for this claim is that the same commitments that prompt Goldman to adopt *historical* process reliabilism should prompt us to adopt the extendedness hypothesis. Goldberg uses the notion of "epistemic reliance" to make this point (2010, 79).[56] When we are evaluating justification, we care not only about what happens at the moment of belief formation, but also the reliability of all of the processes on which the belief is epistemically reliant. Again, consider the case of Stephen above or instances of beliefs based on memories: If I base my belief that p on a memory of having seen that p, we care not only about the reliability of my memory recall, but the reliability of the perceptual faculties that originally caused me to believe that p. Simply put, Goldberg's argument is that our dependence on other people in cases of testimonial exchange are similarly cases of epistemic reliance – not epistemic reliance on one's own temporally-prior cognitive processes, but epistemic reliance nonetheless. Where there is epistemic reliance, there is activity that is relevant to the justificatory status of a subject's belief. For the reliabilist, this means that justification-conferring processes track epistemic reliance. As Goldberg puts it:

That said, I am not the first to propose extending justification-conferring processes beyond the cognition of individual epistemic subjects, as we will see in this discussion of Goldberg 2010.

[56] This is very similar to the notion of "epistemic dependence" discussed in section 2.

> "We can then employ the notion of epistemic reliance to formulate the core insight of Goldman's Historical Reliabilism, as follows: reliabilist epistemic assessment is assessment of *all* the processes on which the subject epistemically relies; in these cases, the *process extends as far as the subject's epistemic reliance extends*." (Goldberg 2010, 92)

Insofar as cases of testimony involve subjects epistemically relying on one another, justification-conferring processes extend to account for that reliance, and as such include the implicated cognitive activity of testifiers.

Goldberg is arguing for his extendedness hypothesis on the grounds that it is necessary given the kind of epistemic reliance at play in cases of testimony. I am arguing for the existence of extended interactive justification-conferring processes on the grounds that it is necessary given the kind of epistemic reliance at play in cases of dialogical deliberation.[57] The kind of epistemic reliance we demonstrate in dialogical deliberation is meaningfully different from the epistemic reliance characteristic of straightforward cases of testimony, where this latter kind of case is conceived of as the preservation of content from one interlocutor to another.[58] In straightforward cases of testimony, we epistemically rely on the *reliability* of interlocutors' temporally prior cognitive processes. We rely on testifiers the same way we rely on thermometers; the fact that testifiers are cognizers plays no special or distinct role. In contrast, in dialogical deliberation, we rely on our interlocutors' participation to

[57] In this section I am arguing that, against the backdrop of Mercier and Sperber's interactionist theory, acceptance of Goldberg's extendedness hypothesis should prompt the acceptance of extended interactive belief-forming processes. That said, I want to note that there are plausible reasons one might accept my contention that justification-conferring processes extend beyond individual cognizers in cases of dialogical deliberation but reject that the same is true in straightforward cases of testimony. In particular, in straightforward cases of testimony (in which the testifier asserts that *p* and the recipient of the testimony comes to believe that *p*), one could argue that the epistemic reliance at play can be completely justificatorily accounted for by evaluating the reliability of the testimony-recipient's judgement that their interlocutor is trustworthy (in the relevant ways). A similar explanation cannot be given in interactionist cases of dialogical deliberation. In such cases, it is not merely interlocutors' judgements of one another's reliability that yield epistemic gains, but rather the triggering or prompting effects interlocutors' contributions have on one another's faculties of reasoning. Therefore, one cannot account for the epistemically relevant history of a deliberative participant's belief solely by examining their individual, isolated cognitive activity.

[58] One might argue that the epistemic relationship I am arguing for bears a resemblance to the notion of "epistemically engineered environments" that Goldberg develops in other work (2020, 2783). Goldberg characterizes this context as "an environment that has been deliberately designed so as to decrease the cognitive burden on individual subjects in the attempts to acquire knowledge" (2795). Although it is certainly the case the Goldberg is investigating a kind of social epistemic reliance, again his focus is not on the kind of epistemic reliance I have in mind. He is interested in a kind of reliance on others' epistemic good will, which is closely related to reliance on other's reliability – in particular, reliance on others to reliably set up environments to ensure or promote our epistemic success.

trigger in ourselves the kinds of cognitive activity that, when deployed in concert with others, allows a group to collaborate its way to a well-reasoned judgment. In such cases, an interlocutor epistemically relies on their deliberative peers to engage in such a manner that reason-giving and reason-evaluating cognitive capacities are tripped. The individuals' reason-giving and reason-evaluating capabilities are not necessarily themselves reliable but contribute to reliability when deployed in dialogue with the joint cognitive activity of others.

To reiterate: importantly, and in contrast with Goldberg's notion of epistemic reliance, this interactionist notion of reliance is not a kind of dependence on one another's reliability. Interlocutors are more than mere thermometers. In light of the interactionist theory, we rely on our interlocutor's dissent, or call for justification, even if that dissent is not itself justified, or the call for justification is unwarranted in some way.[59] This is, of course, not to say that epistemic reliance on reliability is not a genuine and pervasive epistemic phenomenon. It is just to say that there is another, important kind of epistemic reliance; let's call it interactionist epistemic reliance.

4.5 Summary: Justification Tracks Epistemic Reliance

When a subject believes the conclusion of an argument, whether or not she is justified in believing the conclusion depends not only on the epistemic permissibility of her inference, but her justification for believing each of the premises. She epistemically relies on historically prior epistemic events (that confer justification on her beliefs in the premises), and as such those prior epistemic

[59] Neil Levy and Mark Alfano make a related point (2020). However, although they are arguing that we can gain "knowledge from vice", they are not defending a kind epistemic reliance distinct from dependence on interlocutor reliability (1). Rather, their contention is that that "some of our most significant epistemic achievements" result from behavior that looks epistemically vicious. Levy and Alfano's key examples are of intergenerational epistemic success, primarily cases in which humans are able to successfully navigate and flourish in challenging and diverse environments. They argue that these kinds of ecological epistemic successes result from an innate human disposition that is, at the individual level, an epistemic vice: extreme overimitation (7, 11). These cases are incredibly interesting and should prompt any reliabilist interested in social epistemology to think about how we ought to temporally carve up processes. Is the relevant justification-conferring process the short, immediate, local process? Or is the relevant justification-conferring process extremely temporally extended (even intergenerational)? Ultimately however, Levy and Alfano are arguing that behavior the virtue epistemologist would call vicious can be reliable, not that there is a kind of epistemic reliance distinct from dependence on interlocutor reliability.

events are part of what confers justification on her belief in the conclusion of her inference. The relevant justification-conferring process is *temporally* extended insofar as the subject is epistemically relying on historically prior processes. If we follow Goldberg, in straightforward cases of testimonial exchange, the recipient of the testimony epistemically relies on the testifier in forming a belief that p when the testifier asserts that p. The relevant justification-conferring process is *temporally* and *interpersonally* extended insofar as the subject is relying on a historically prior process that occurs in another epistemic subject. In dialogical deliberation, interlocutors epistemically rely on one another to trip certain cognitive, reason-giving and reason-evaluating capacities. Interlocutors depend on one another to critically engage, to divide cognitive labor, so that the group can collectively navigate its way to the truth. When an interlocutor forms a belief as the result of social deliberation, the relevant justification conferring process is *temporally* extended and *interactively* extended.[60]

5. Deliberative Dialogue and the Case Against Evidentialism

In the previous section, I demonstrated how process reliabilists, in their commitment to the historical dimension of justification, can easily incorporate the epistemic consequences of Mercier and Sperber's interactionist theory of reasoning. At this juncture, it is appropriate to question whether other theories of justification can make similar accommodations: can other theories satisfy the explanatory desideratum if Mercier and Sperber's theory is true? I will argue that other prominent

[60] The extended interactive justification-conferring processes I defend here bear a resemblance to the "social belief-forming processes" Joseph Shieber proposes (2019). Our views converge insofar as we are both arguing that "the belief forming process that results in a believer's belief need not be limited to the cognitive processes of the believer alone" (93). However, there are important differences between my view and Shieber's, both in terms of how we are conceiving of these social processes, and our arguments in their defense. It is not clear that Shieber would endorse my proposal that processes can extend interactively to account for the reliability of deliberative dialogue (as opposed to merely the aggregate reliability of individual cognizers). Shieber argues (using both thought experiments and empirical data) that we should be suspicious that the justification of beliefs formed on the basis of testimony can be explained by our competence for assessing reliability. This, he argues, is particularly clear in cases where the testimony in question is the result of a "socially distributed systems of information transmission", largely because of the opacity of such systems as well as the "non-locality of expertise" (90). In light of this, it seems like he is primarily concerned with social epistemic reliance on reliability, not the kind of interactive epistemic reliance I discuss here. It seems that insofar as Shieber casts his discussion in terms of testimony, he is primarily concerned with accounting for the way we learn *from* others in those cases where the testifier is socially distributed, while I am primarily concerned with accounting for the way we learn *with* others.

theories, namely internalism and specifically evidentialism, cannot. As I outlined in section 2, the commitment to the historical dimension of justification and the explanatory desideratum are credible and widely held positions. In this section, I argue that Mercier and Sperber's interactionist theory of reasoning shows us how demanding those commitments can be and contend that internalism is unable to meet the corresponding demands.

Let us start by considering what resources are available to the access internalist. Access internalism is the view that justification is a function of whether a subject properly bases her belief on, or whether a subject's belief properly coheres with, other beliefs to which she has conscious, cognitive access (Bonjour 1980). An access internalist may try and argue that they can account for the justificatory status of beliefs formed during dialogical deliberation by taking stock of the reasons and arguments deliberative participants consciously possess at the end of the deliberation. Returning to the example of the hiring committee discussed above, the access internalist might argue that at the end of their discussion, Madeleine, Tanja, and Kevin not only have a belief that Jesse is the most-qualified candidate, but they also possess a consciously accessible argument in favor of Jesse's candidacy. After all, the back-and-forth of discussion has produced publicly articulated reasons and arguments, as well as evaluations of those reasons and arguments. The access internalist might say that it is interesting to acknowledge that these were generated in a social, dialogical fashion, but the justificatory status of the belief that Jesse is the most qualified is conferred by the committee members basing this belief on whatever reasons they (consciously) possess at the end of the debate.

This analysis, however, is not plausible. Mercier and Sperber's argument is not that conscious, reflective reasoning works well, and evolved for use in, conversation with others because it allows us to generate consciously accessible arguments in defense of the positions being discussed.[61] No part of

[61] See footnote 24 for a discussion of this point is relevant to my externalist reliabilists analysis of Mercier and Sperber's interactionist theory.

their view commits us to the idea that deliberative participants are able to recite the arguments considered during the course of the discussion after the conversation has concluded.[62] Even if there are cases where interlocutors possess consciously accessible arguments in favor of the beliefs they form as the result of deliberation, we ought not think (as the access internalist would like us to) that it is the arguments' conscious accessibility that is conferring justification. We must recognize that on Mercier and Sperber's view, the epistemic gains of dialogical deliberation are not a function of the complete, consciously held well-reasoned arguments interlocutors possess at the end of discussion. Rather, the epistemic gains are the result of the dialogical interaction – the way in which dialogue *prompts* certain epistemically useful cognitive processes. This illuminates the more pressing concern with this access internalist's explanation. Like the individualist reliabilist strategies rejected above, this access internalist approach fails to satisfy the explanatory desideratum; it offers an incomplete explanation for why participants in sufficiently cooperative and collaborative deliberative dialogues non-accidentally form true beliefs. By the lights of the interactionist theory, the pieces of reasoning that are considered "bad" when evaluated in isolation (e.g., biased reasoning) are part of the explanation for why the beliefs formed in such settings are non-accidentally true. As such, these cognitive events are in some respect justification-conferring. Given the individual reasoning that is

[62] It is worth noting that this consideration also highlights that the access internalist is vulnerable to Goldman's argument from "preservative memory" against time-slice theories of justification (Goldman 2001, 323). Preservative memory refers to our cognitive capacity to store beliefs we've formed over extended periods of time and then recall them without also recalling their evidential bases, all the while preserving the justifiedness of the belief. There are all sorts of facts we *know* even though we do not recall the evidential source that gave rise to the relevant belief, or the arguments we originally possessed in its defense. Despite not remembering the source, insofar as knowledge entails justification, we are justified in believing these propositions. This prompts Goldman to conclude that "[i]f *S* has a justified attitude *D* towards proposition *P* at *t*, and is *S* retains attitude *D* towards *P* until the later time *t'*, via memory, then, *ceteris paribus*, *S* is still permitted to have attitude *D* towards *P* at *t'*" (Goldman 2001, 323). Given that events prior to moment of belief evaluation are relevant to the belief's justificatory status this demonstrates that a historical account of justification is needed. How does this relate to our discussion of the interactionist theory of reasoning? Even if we grant that Tanja, Kevin, and Madeleine formed consciously accessible arguments at the end of the discussion, Goldman will likely argue it is not the possession of these arguments that is doing the justificatory work given it would be appropriate to say the following: at some time in the future, after they had forgotten those arguments, they are *still* justified in believing that Jesse is the most qualified candidate.

"bad" when considered in isolation plays no role in the access internalist's justificatory story, the access internalist falls short of accommodating the interactionist theory.

What about mentalist internalists? Do they fare better? The mentalist internalist argues that the justificatory status of a subject's belief supervenes on the subject's mental states at the moment of belief evaluation. Importantly, mentalist internalists are traditionally time-slice theorists. To figure out the justificatory status of a subject's belief, we look that the subject's mental states *at the time of belief evaluation*. Consider Conee and Feldman's characterization of evidentialism, a species of mentalist internalism: "Doxastic attitude D toward proposition p is epistemically justified for S at t if and only if having D toward p fits the evidence S has at t" (Conee and Feldman 1985, 15; underline mine). Given the theory posits that a belief's justificatory status can be determined by taking a "snapshot" of the subject's mental states at the moment of belief evaluation, this position is incompatible with accepting that justification has an integral historical dimension. As such, like access internalism, traditional evidentialism will leave the pieces of reasoning that are considered "bad" when evaluated in isolation out of the justificatory story in cases where beliefs are the result of deliberative dialogue. Consequently, evidentialism similarly fails to satisfy the explanatory desideratum.

That said, internalist evidentialists can resist the time-slice characterization of their view and argue that the fundamental insights of the theory can be preserved while adopting a commitment to the historical dimension of justification. For example, one could argue that evidentialism requires that subjects properly base their beliefs on their evidence, and there is a "diachronic requirement" on proper basing (Fantl 2020: 784-786). The central idea is that proper basing takes time; in order for your belief that p to be properly based on your evidence e, then e has to cause you to believe (or cause you to continue to believe) that *p*. So long as the evidentialist has an argument for the claim that epistemic causes always precede their effects, they can reject time-slice epistemology. Does *historical*

evidentialism, as species of mentalist internalism, have the resources to accommodate the epistemic consequences of Mercier and Sperber's interactionist account?

It does not. Seeing this requires taking a step back and looking more broadly at the criticisms I have been making of individualist and internalist approaches to handling the epistemic consequences of Mercier and Sperber's interactionist theory. Again: evaluations of justification are ultimately a matter of figuring out what makes it the case that the subject's belief is non-accidentally true. Paired with a commitment to the historical dimension of justification, the epistemic consequences of the interactionist theory demonstrate that part of what explains that beliefs formed in (epistemically healthy) social, deliberative ways are segments of individual participants' reasoning that are "bad" when considered on their own, separate from the rest of the interaction. Therefore, (at least part) of what is justification conferring in these cases is not evidential. Fundamentally, this is what internalism fails to appreciate, and it is what renders it incapable of satisfying the explanatory desideratum for theories of doxastic justification.

The historical evidentialist might try and insist that the role of this "bad" reasoning is in some sense accounted for: after all, the "bad" reasoning prompts the production of the good evidence, and evidence is at the heart of the justificatory story on their view. However, this response will not do. Consider two groups, one that engages in collaborative, cooperative reliable dialogue, another that engages in unreliable dialogue (perhaps as a result of there not being sufficient disagreement between group members, which can lead to groupthink). By chance, the unreliable group happens to get lucky on this occasion. Despite not engaging collaboratively and cooperatively with one another and interacting in a manner that generally leads to poor epistemic outcomes, at the end of their discussion the deliberative participants base their true beliefs on the same evidence as the group that reasoned reliably. Given the two groups end up with the same evidence, the historical evidentialist can't account for the intuition that the latter, but not the former, group has justified beliefs at the end of their

deliberation. On the framework of the interactionist theory, these cases demonstrate the limits of internalist theories to account for the historical dimension of justification. This is because, as I have shown, the important epistemic consequence of Mercier and Sperber's work is that not all justificatorily relevant historical factors cannot be accounted for evidentially.

6. Anticipating Objections

In this section, I anticipate and respond to potential objections. The first objection is that my argument is at odds with justification's status as an evaluation of how individuals epistemically handle their environments. The second objection contends that deliberative dialogue is better thought of as a context in which a process is used rather than as a constituent part of a justification-conferring process. The third objection charges my argument with conflating the distinction between what causes a subject to form a belief and what confers justification on a belief. I will take each objection in turn.

6.1 Epistemic Credit

When originally formulating process reliabilism in "What Is Justified Belief?", Goldman suggests we "restrict the extent of belief-forming processes to '*cognitive*' events, i.e. events within the organism's nervous system" given justification is an evaluation of "how a cognizer deals with his environmental inputs" (1979, 13). The thought is that reliabilists can best capture the way in which justification is a measure of an individuals' epistemic competence by limiting belief-forming processes to individuals' cognitive systems. Those committed to this idea might reject my defense of extended interactive belief-forming processes on the grounds that it undermines this intuition. In response, I argue that Mercier and Sperber's theory entails that, often, cognizers deal with their environmental inputs by working in collaboration with others. Insofar as this is the case, extended interactive belief-forming processes seem totally appropriate.

However, I believe that those who continue to press this objection may be motivated by another intuition: the idea that justification is a matter of whether or not an individual deserves

epistemic credit. Some will argue that when a subject knows that *p*, they deserve credit for believing that *p* (see for example, Greco 2003, 123). Proponents of this view may argue that insofar as extended interactive belief-forming processes include the cognitive activity of interlocutors, I am implausibly suggesting deliberative participants get undue epistemic credit for their interlocutors' epistemic work. In response, I would say that it is very difficult to maintain this position while taking social epistemology seriously. Social epistemologists highlight for us the ways in which we epistemically rely on others. Consider Jennifer Lackey's (2006) argument for understanding instances of testimonial exchange as paradigmatic cases of knowing that *p* without deserving (total) epistemic credit. Often, when a subject permissibly forms a true belief on the basis of an interlocutor's testimony, the bulk of the explanation for why the subject's belief is non-accidentally true involves "the cognitive resources of someone other than the subject in question" (Lackey 2006, 356). If we are comfortable claiming that subjects are justified when they form beliefs on the basis of their interlocutors' testimony despite not deserving full credit for the non-accidentality of the belief in question, then we should reject the idea that knowledge and justification are matter of epistemic credit. Once we reject the stipulation that justification and knowledge are matters of epistemic credit, we can more comfortably accept that justification-conferring processes in some cases include the cognitive activity of subjects other than the one whose belief is under evaluation.[63]

6.2 Contexts versus Processes

Many reliabilists will plausibly argue that when evaluating justification, we look at the reliability of a belief-forming process relative to a particular context. On this view, whether a belief is justified depends on whether the relevant token belief-forming process is "reliable in environments of the same

[63] Alternatively, we might grant that justification is a matter of epistemic credit and argue that Mercier and Sperber's interactionist theory undermines the stronger claim that justification for a subject's belief that *p* is a matter of whether that solitary subject deserves credit. Rather, epistemic credit can be distributed amongst interlocutors.

types as the one in which the belief was formed" (Heller 1995, 504).[64] Those sympathetic to this view may contend that dialogical deliberation is not a constituent part of a justification-conferring process, rather it is a context in which a justification-conferring process can take place – namely, a context in which conscious, reflective reasoning is reliable according to the interactionist theory. One might argue that to include dialogical deliberation as part of a justification-conferring process would be akin to including the quality of light, or the functioning of a light source, as part of the process that confers justification on a visual perceptual belief. The quality of light, or the functioning of a light source, is clearly not part of a process that confers justification. It is true that in such cases the functioning of a light source is part of the history of belief formation, but many will argue that the appropriate way to account for this is to claim that the reliability of a cognitive process that confers justification is determined relative to the context of use. In other words, (human) visual perceptual processes are not reliable *simpliciter*, but rather reliable in contexts with sufficient/adequate light. One might object to my argument above by claiming that, similarly, the reliability of certain cognitive capacities is determined relative to the social features of one's context. To oversimplify, the idea would be something like the following: conscious, reflective reasoning capacities are reliable in contexts with sufficient amounts of deliberative, interactive dialogue, and unreliable in contexts in which individuals are isolated.

This position is prima facie plausible and maintains the conservative individualist intuition that justification-conferring processes are confined to individuals' cognitive systems. Why argue for the more radical conclusion that there are extended interactive justification-conferring processes if it is

[64] Mark Heller defends this idea in responding to the generality problem for reliabilism – the objection that there is no principled level of specificity for individuating belief-forming process types when evaluating reliability. Heller argues that the generality problem is not a substantive objection, so long we take a contextualist approach to evaluations of reliability. Goldman (1979, 86-91) discusses the idea that the reliability of process-types are judged relative to particular contexts of use in crafting a relevant-alternatives response to the well-known barn facades case. In a county full of facsimile barn façades, when one forms the visual perceptual belief that one is looking at a real barn, a relevant alternative is that one is looking at a barn façade. One could argue that given this relevant alternative, this context in not one in which the relevant visual belief-forming process is reliable.

not necessary? However, this dialogue-as-context position doesn't sufficiently capture the phenomenon as Mercier and Sperber have described it. Recall the discussion from section 3.2. Mercier and Sperber's claim is not that individuals' reliable solitary reflective capacities are tripped when engaging in dialogical deliberation. Rather, their claim is that our capacity to defend our beliefs, to produce reasons in support of our beliefs, and to evaluate interlocutors' arguments are tripped in dialogic deliberation. These competences are not individually truth-tracking. Rather, deployed collectively, they help deliberating groups reach the truth. Insofar as we don't become individually more reliable in dialogical contexts, we can't accommodate the interactionist theory by positing that deliberative dialogue is a context relative to which individual conscious reflective reasoning must be evaluated.

6.3 Causation versus Justification

Some might object to my defense of extended interactive belief-forming processes by asserting that even a proponent of historical accounts of justification is not committed to saying that everything in a belief's causal history is epistemically relevant – that is, justification-conferring or justification-detracting. For example, one could plausibly argue that both my drinking a cup of coffee this morning and the processing of my visual perceptual faculties are causally related to my present justified belief that my cat is asleep in the window. My cognitive processes would undoubtedly have functioned differently had I not consumed coffee this morning, likely more sluggish, and therefore my drinking coffee this morning had some causal impact on my belief about the cat's whereabouts. Despite this causal connection, most proponents of historical accounts of justification plausibly think it's inappropriate to say that coffee consumption is part of the process that confers justification on my belief. This example helps illustrate the important distinction between that which *causes* a belief, and that which *confers justification* on a belief.

Just as my coffee consumption is justificatorily irrelevant to my belief about my cat's whereabouts, one might argue that deliberative dialogue is similarly justificatorily irrelevant to beliefs I form in discussion with others. Deliberative dialogue may be part of a belief's causal history, but that doesn't mean it is necessarily justificatorily relevant. If we ought not think of deliberation and interaction as epistemically evaluable parts of the causal histories of beliefs, then, of course, we should abandon the idea that these deliberations and interactions should be thought of as part of extended interactive belief-forming processes that confer justification.[65]

In responding to this objection, it is important to think about what explains the intuition that drinking coffee is not part of an epistemically evaluable process, despite the effect it may have on an epistemically evaluable process. Consider our visual perceptual processes, often taken to be the reliabilist's paradigmatic example of a justification-conferring process. It would be helpful to think about why visual perceptual processes are epistemically evaluable or justification conferring, with the aim of identifying how drinking coffee is relevantly dissimilar. Here is one explanation: It is appropriate to epistemically evaluate visual perceptual processes, and understand them as justification-conferring, because our visual perceptual systems evolved to help us form accurate beliefs about the world around us. Accurate belief formation is what our visual cognitive processes are for; the teleological role of our visual perceptual faculty is distinctly epistemic. In contrast, the way in which coffee is metabolized does not have the same distinctly epistemic teleological role of improving the reliability of cognition. It is doubtful it was selected for any such role. Rather drinking and metabolizing coffee, like metabolizing other sources of sustenance, is a process of energy conversion that supports general metabolic function. It serves many functions in addition to epistemic cognitive

[65] It might be helpful to clarify the distinction between this objection and the objection discussed in the previous section. The previous objection granted that dialogical deliberation can be justificatorily relevant but rejected cashing-out the epistemic relevance of deliberative dialogue in terms of justification-conferring processes. In contrast, this objection takes issue with the idea that dialogical deliberation is relevant to evaluations of justification.

functioning. Insofar as coffee consumption is lacking a distinctly teleological-epistemic watermark, it cannot be said to be part of, or a distinct, justification-conferring process.

Given this analysis, what should we say about the social dialogical interactions that are the focus of this paper? Do they have the required teleological-epistemic watermark necessary for a process to be evaluated as justification conferring? The answer is yes. As is discussed above, Mercier and Sperber's defense of their interactionist theory of reasoning is a teleological analysis. Their argument proceeds by looking at the empirical data about our ability to reason, and then formulating an explanation of that data that makes evolutionary sense insofar as it characterizes reasoning as a *successful* function. On examining the data, what they determine is that the best way to understand conscious, reflective reasoning as properly adaptive cognitive functioning is as part of our social epistemic lives. Conscious reflective reasoning is instrumental to epistemic success when deployed in interactive engagement with others. The claim is that conscious, reflective reasoning is adapted for use in deliberative interaction because that is when it leads to epistemic success.

Recall that Mercier and Sperber's interactionist theory of reasoning is proposed as a replacement for the classical intellectualist theory of reasoning. The intellectualist theory posits that conscious reflective reasoning is a cognitive function with an epistemic teleological watermark, but one that is designed for individuals to deploy in isolation. Mercier and Sperber's central criticism of this view is that, given the epistemic errors that conscious reflective reasoning exemplifies when deployed in isolation, it doesn't make sense to identify isolated, individual epistemic gain as the adaptive function of this kind of cognition. The central advantage of the interactionist theory of reasoning is that it gives an explanation of conscious reflective reasoning that is compatible with the empirical data on when this kind of reasoning reliably leads to epistemic success. The result is that conscious reflective reasoning in social dialogical interaction can be said to be an adaptive ability with a distinctly epistemic function in the same sense that our visual perceptual processes have a distinctly

epistemic function. As such, it would be misguided to classify interactive dialogue and deliberation with non-epistemic causal influences on our belief forming processes.

One other reply to this formidable objection is necessary. In addition to the discussion above, it is important to note that coffee-drinking is part of the causal history of belief-formation insofar as it has an impact on the reliability of an *individual* cognizer's cognition. When we are insufficiently caffeinated or sleep-deprived, we become, as individuals, less reliable. To the extent that anyone would want to account epistemically for coffee-drinking, it seems that it would be appropriate for them to do so by discussing contexts or environments in which processes occur. Sleep deprivation may result in blurry vision, but we don't want to say that, as a result, visual perpetual processes are unreliable and can't yield justified beliefs. Rather, we should say that visual perceptual processes are reliable relative to contexts with sufficient lighting and when carried about by a sufficiently alert human subject. But recall from the discussion in section 4.2, insofar as dialogical deliberation doesn't make individual reasoning more reliable, we can't accommodate the interactionist theory by similarly discussing proper contexts of use.

7. Conclusion

I have argued that if Mercier and Sperber's interactionist theory is true, process reliabilist should accept that extended interactive justification-conferring processes follow from their plausible commitment to historical theories of justification. Moreover, I have demonstrated that an interesting consequence of the interactionist theory is that not all parts of the justification story can be cashed out in evidential terms. By way of conclusion, I want to note that over and above accommodating the insight that we epistemically rely on one another's deliberative engagement, extended-interactive belief-forming processes give us the tools for exploring and understanding interesting social epistemic phenomena in a relatively conservative fashion. When social epistemic feats and vices call for theorizing, one option is to posit collective epistemic subjects with their own cognitive systems that

can be the loci of epistemic evaluation. While offering this kind of explanation is an understandable impulse, for many it is too metaphysically radical. Do groups really have beliefs? Are they really subjects of epistemic evaluation? Or are we merely speaking in metaphor and generalization when we say that groups have (un)justified beliefs? By positing extended interactive justification-conferring processes, we epistemically account for role of group discussion and collaboration without taking on the metaphysical burden of defending group belief and justification. This is, in large part, because the view has a far more expansive notion of what can confer justification than internalist and traditional externalist views.

Moreover, extended interactive justification-conferring processes point to interesting explanations of social epistemic phenomena like echo chambers and epistemic bubbles. Epistemic bubbles are environments in which certain viewpoints and voices are absent, and echo chambers are environments in which certain viewpoints are "actively excluded" and "discredited" (Nguyen 2020, 1). In so far as interactionist reasoning thrives on disagreement, the theory predicts that reasoning may lead to epistemic distortions in environments where disagreement is absent or hindered. By acknowledging that justification-conferring processes can be extended interactively, we can explain why subjects in echo chambers and epistemic bubbles end up saddled with epistemically distorted beliefs. This is, of course, just a sketch of the kind of interesting work that could be pursued should we accept extended interactive justification-conferring processes. However, my hope is that it highlights that my view can provide deeply social epistemic explanations without burdensome metaphysical commitments.

CHAPTER 3

EPISTEMOLOGY'S BLAME GAME

1. Introduction

Let's talk about the new evil demon in town: echo chambers. We all live in, and rely heavily on, our epistemic communities. There is a network of sources that we trust, and sources that we distrust. In a healthy epistemic community, those trust and distrust relations are reliable; they help us form true beliefs and avoid forming false ones. Unfortunately, our epistemic communities can be corrupted in various ways. In an echo chamber, trust relations are such that an epistemic community becomes problematically insular and isolated. Specifically, and following Thi Nguyen, an echo chamber is "a social epistemic structure in which other relevant voices have been actively discredited" (Nguyen 2020, 2). Common examples of pernicious echo chambers are anti-vaxer and climate change denier communities. Nguyen argues that these groups have the following features. First, these epistemic communities "exclude non-members through epistemic discrediting" (6). Second, membership in such communities requires accepting a "core set of beliefs", which includes beliefs that sources external to the community are unreliable (6). Together, these features result in an epistemic community in which there is "significant disparity in trust between members and non-members" (6). Why can this be a problem? Once a false belief is in an echo chamber, it will proliferate within the community and become difficult to dislodge, particularly if it is a core belief that is "prerequisite" for community membership (6). Insofar as people in echo chambers trust community members, and not sources outside the community, it becomes

incredibly difficult for reliable outside sources to penetrate the community and correct epistemic errors.[66]

Importantly, echo chambers are a distortion of the necessary, indispensable work that epistemic communities do. As Nguyen writes:

> It is important to note that the epistemic mechanisms by which echo chambers work, though problematic, are not sui generis. They are perversions of natural, useful, and necessary attitudes of individual and institutional trust. (8)

Trusting some, distrusting others, and using community leaders as a means of figuring out how to navigate trust relations are all indispensable and significant parts of our epistemic lives. Just like those trapped in an echo chamber, subjects in healthy epistemic communities have a standing trust in other members of their community. They are reasonably less confident in the credibility of those outside their community. And importantly, they use their community members (in particular community leaders) to navigate the trust they ought to place in external information sources. These considerations lead us towards the following frightening possibility: epistemic subjects can't be sure, from their own perspective, whether they are in "the bad case" (a pernicious echo chamber) or "the good case" (a healthy epistemic community). The more we learn about echo chambers, the more it seems like the following scenario is possible: There are two epistemic subjects, one born into an epistemically healthy community, another born into an echo chamber. Other than this, the subjects are epistemically

[66] Dramatically, confronting (even a reliable) outside source that offers testimony contrary to the beliefs characteristic of the echo chamber can prompt members to become *more* distrustful of that outside source, and consequently more insular. For example, distrust of outside sources can be cultivated through what Nguyen calls *disagreement-reinforcement mechanisms* (2020, 7). Members might be told, perhaps by a community leader, to anticipate that outside sources will lie and claim that *p*, thus preemptively discrediting the testimony of outside sources. Moreover, in such cases echo chambers' receipt (and rejection of) the testimony of the outside sources can result in a boost in credibility for the community leader: the community leader did, after all, accurately warn that outside sources would claim that *p*. Like voices in a cave, the beliefs within and echo chamber reverberate amongst community members. They hear and believe one another, actively discredit and distrust outsiders. As such, epistemic errors in an echo chamber are (i) difficult for community members themselves to identify and (ii), difficult to correct.

similar. They have the same epistemic dispositions, virtues, vices, and instincts. However, the subject born in the echo chamber has lots of problematic false beliefs, while their counterpart in the healthy epistemic community does not.

Echo chambers seem analogous to familiar cases from the epistemological literature. Being in a pernicious echo chamber involves unknowingly being in the "bad case", just like the person in Stewart Cohen's New Evil Demon problem (1984) or the brain in a vat.[67] Echo chambers are analogous, but arguably more frightening; the possibility that we are trapped in a pernicious echo chamber seems more likely than the possibility that we are deceived by demons or brains in vats. All to say, consideration of echo chambers brings a certain urgency to epistemological theorizing.

To that end, this paper aims to investigate an epistemological idea that is often called on to explain cases like this: epistemic blame. The pull towards an explanation that appeals to the notion of epistemic blame is powerful. After all, it is not the person's fault that they are being deceived by a demon, that they are a brain in a vat, that they are trapped in a pernicious echo chamber. Therefore, they must be epistemically blameless.

This line of reasoning is drawn on by both epistemologists who are internalists and externalists with respect to epistemic justification. For internalists, this intuition about epistemic blamelessness is evidence that what is epistemically significant is internal to epistemic agents (e.g., Cohen 1984, 281). The externalist can say that while the subject in the echo chamber has lots of problematic unjustified false beliefs (say, because the belief-forming

[67] In Cohen's New Evil Demon problem we are asked to imagine two subjects who are in two different worlds (with different physical contents), but who have internally identical qualitative perceptual experiences. One subject is a world where the perceptual experiences are veridical and reliably caused by physical objects in their world. The other subject's perceptual experiences are non-veridical in their world and are being caused by an evil demon rather than by the physical objects in their world. Cohen's intuition is that such a case shows that reliability isn't necessary for doxastic justification. After all, Cohen argues, insofar two subjects are having identical perceptual experiences it would be implausible to claim that their beliefs have different justificatory statuses.

processes that are characteristic of echo chambers are unreliable), she is epistemically blameless with respect to those beliefs.[68] She is unjustified, but it's not her fault. Indeed, there is a view in the literature that externalists need a notion of epistemic blame for their theory to be plausible.[69] The idea is that without it, they have no way of responding to the objection that their view gives implausible verdicts.[70] Appealing to epistemic blame allow externalists to accommodate internalist intuitions while maintaining their theory of doxastic justification.

Despite this, there has been relatively little investigation into the relationship between theories of doxastic justification and epistemic blame.[71] This chapter aims to fill that gap. In particular, I am interested in exploring how an epistemologists' internalist or externalist commitments regarding the nature of doxastic justification influence or constrain their theorizing about epistemic blame. I will start by disentangling two distinct approaches one could take when theorizing about epistemic blame. In short, one could investigate epistemic blameworthiness, or epistemic blaming. I then turn to the project of thinking about how internalists and externalists with regards to doxastic justification can make room for a notion of epistemic blame on their frameworks. I start with internalism, and argue that there is no

[68] E.g., Brown (2020a and 2020b), Srinivasan (2020), Hawthorne and Srinivasan (2013).

[69] For defenses and discussions of this kind of view, see: Brown (2020a, 2020b); Boult (2017a, 2017b, 2019, 2021b); Williamson (forthcoming); Littlejohn (forthcoming); Srinivasan and Hawthorne (2013); Srinivasan 2020; and Ballarini 2022.

[70] Boult concisely summarizes this attitude: "[G]iven that radical externalists maintain that epistemic norms are factive, it seems there can be pairs of cases in which agents are unjustified with respect to a factive norm, but where there is nevertheless an intuitive normative difference between the cases. Perhaps agent A formed a false belief by wishful thinking, and agent B was carefully deceived. A and B both have unjustified beliefs by radical externalist lights; given the intuitive normative difference between the cases, we might worry whether radical externalists have a correct understanding of the nature of justified belief. In order to respond to this sort of worry, radical externalists typically appeal to the idea that the difference here comes down to a difference in blamelessness/blameworthiness. Agent A is unjustified and blameworthy, while agent B is unjustified but blameless. Similar points can be made about internalist evidentialism and views that fit somewhere between these two extremes, such as simple forms of reliabilism" (2021b, 3).

[71] Hawthorne and Srinivasan discuss the possibility of having two distinct kinds of epistemic norms, a subjective 'ought' (that tracks praise and blame) and an objective 'ought' (that tracks epistemic outcomes (2013). This is cashed out in a more general discussion about the nature of epistemic normativity, not doxastic justification. They are pessimistic about this approach, which they term a "two-state solution" (2013, 25-28).

room for a notion of epistemic blameworthiness as distinct from doxastic justification on the internalist framework. While some argue that this just means that epistemic blaming is the interesting and novel investigation internalists can pursue, I argue that we shouldn't expect conceptual analyses of epistemic blaming to be philosophically illuminating. I then turn to examining to what extent the externalist framework can accommodate a notion of epistemic blame. By exploring different evaluative pairings[72], I argue that the externalist can't endorse the required norm of epistemic responsibility. However, I argue that externalists should be fine with this result. Far from needing a theory of epistemic blame to make externalism plausible, rejecting a robust notion of epistemic blame that tracks internalist intuitions should be understood as a feature, not a drawback, of their view.

2. Two Distinct Approaches: Epistemic Blameworthiness vs. Epistemic Blaming

There are two approaches we can take when theorizing about epistemic blame. One approach is to conceive of the relevant philosophical project as a normative investigation into whether belief states have particular properties such that the subject is blameworthy with respect to those states. In contrast, one could conceive of the relevant philosophical project as a primarily descriptive investigation into what constitutes a genuine act of blaming. These two approaches are distinct insofar as they are ultimately concerned with different questions. That said, they are certainly compatible; one could be interested in both questions. In this section I want to clearly disentangle the two projects. My concern is that if we don't, epistemologists interested in epistemic blame will end up talking past one another.

[72] Blamelessly unjustified, blameworthy and unjustified, etc.

2.1 Epistemic Blameworthiness

The first approach takes theorizing about epistemic blame to be a matter of investigating a normative epistemic evaluation. What are the standards or norms relevant to determining whether a person's belief is epistemically blameworthy? What makes it the case that someone is, in fact, epistemically blameworthy? On this view, epistemic blameworthiness is of a kind with epistemic justification. It is criticism of an individual with respect to belief states with particular properties.[73] However, it is distinct from epistemic justification insofar as it tracks distinct epistemic standards. On this approach, the central focus is identifying the standards epistemic blameworthiness tracks. Generally speaking, the idea is that epistemic blameworthiness tracks culpable violations of particular epistemic norms. As such, the relevant philosophical work involves unpacking the necessary and sufficient conditions of a distinctly epistemic kind of culpability. Relevant questions include the following: Are epistemic subjects required to have (some kind of) voluntary control over their beliefs in order to be held to account for possessing them? Do epistemic subjects have to be (consciously) aware of particular epistemic norms in order to be held to account for violating them? Let's call this approach an investigation into epistemic blameworthiness.[74]

2.2 Epistemic Blaming

A distinct approach investigates, and aims to offer a unified account of, the practice of epistemic blaming. When I epistemically blame another person, what is it precisely that I am doing? Does the act of epistemically blaming another person require that the person doing the epistemic blaming feel particular emotions, have particular attitudes, or hold particular beliefs? Does epistemically blaming another require that the person doing the blaming act in a particular way? Let's call this second approach an investigation into epistemic blaming.[75]

To see how an inquiry into epistemic blaming is distinct from an inquiry into epistemic blameworthiness, consider: we can talk about what justification is, i.e., what it takes for a subject's belief to be justified, without discussing the social, communicative practice of asserting that someone's belief is justified.[76] Similarly, the question of what it takes to assert or attribute knowledge is separate from the question of what makes it the case that someone is a

[73] There are interesting questions about whether it is appropriate to conceive of the subject or their belief state as what is being evaluated as blameworthy. I am choosing to conceive of the evaluation as follows: people are epistemically blameworthy with respect to particular belief states because those belief states have particular properties. I anticipate this decision is somewhat contentious. For example, I imagine that some might want to argue that evaluations of epistemic blameworthiness track subjects' epistemic character or stable epistemically vicious dispositions. A proponent of this view may want to argue that while a particular belief might be a manifestation of a subject's bad epistemic character, it is the person qua epistemic subject that is being evaluated. After all, the vice is attributable to the epistemic agent, not the belief state. Ultimately, I don't think we should go this route. Say we grant that it is appropriate to make general epistemic criticisms of a person – i.e., criticisms of a person qua epistemic agent, not merely with respect to a particular belief state. Say we also grant that the notion of epistemic blameworthiness ought to be used to make such a criticism. It is still possible that a person can be blameworthy *merely* with respect to a particular belief state. Consider: it seems completely plausible to say that a person who performs a malicious action intentionally is morally blameworthy even if the action is completely out of character. In short, good people do bad things on purpose sometimes. In such cases, good people are blameworthy with respect to those bad things, even though the bad behaviors aren't expressions of their character or values. Analogously – and assuming that people can be culpable for violating epistemic norms – rational people sometimes reason in culpably epistemically vicious ways. They are epistemically blameworthy when they do, even though the bad instance of reasoning is not representative of their general epistemic character.

[74] Rettler (2018), Brown (2020b), Boult (2021c), and Meehan (2019) engage with this set of questions. However, it is worth noting that this is not precisely how the latter three philosophers frame their projects. Brown and Boult are both primarily concerned with showing that cases in which we blame people for holding particular beliefs, or reasoning in particular ways, can't be reduced to instances of moral or professional blame. Insofar as this is the case, they are aiming to demonstrate that there is a distinct kind of epistemic culpability. However, their focus is on arguing that the norms that are culpably violated in the motivating cases are distinctly epistemic. Neither spends significant time (as Rettler does) giving an account of why it is the case that the epistemic norm violations in the motivating cases can properly be called culpable, i.e., the kind of norm violations for which people can be blamed. Instead, they move from the intuition that something blameworthy has occurred in the relevant case (assuming that a norm has been culpably violated), to an argument for why the norm violation in question is distinctly epistemic (as opposed to moral or professional).

[75] Brown (2020a) and Boult (2021a) engage with this set of questions.

[76] It might seem like Ballarini (2022) draws a similar distinction. However, for Ballarini, the property of blameworthiness has an inextricably social dimension. Blameworthiness is not merely a matter of culpability or responsibility. Rather it is "associated with the appropriateness (or "fittingness") of certain negative attitudes and emotional reactions" (11). Interestingly, the focus isn't on features agent being evaluated or features of the actions she performs. Rather, the focus is on what reactions it would or wouldn't be appropriate for the person doing the evaluating to have. One might argue that these two ways of understanding the property of blameworthiness aren't meaningfully different – or at least, they are two side of the strange coin. After all, one could argue that what reactions it is appropriate for the evaluator to have depends on features of the agent and/or action being evaluated. But I don't think this quite right: the appropriateness of the reaction is a function of whether the beliefs of the evaluator are justified. For example, it may be appropriate for Rory to react with resentment towards Etienne if Rory has very good evidence to believe that Etienne has culpably violated a norm that resulted in harm to Rory, even if it is not in fact the case that Etienne culpably violated the norm.

knower. In that same vein, we should think that the question of what constitutes a genuine act of epistemic blaming is distinct from the question of what it takes to be genuinely epistemically blameworthy.

Comparison with the moral case clarifies this issue as well. We can discuss what makes it the case that a person is blameworthy for their action, which is a matter of explaining why they are culpable or responsible for their action, without discussing whether we ought, all-things-considered, express to that person that they are morally blameworthy. It seems like, in the moral case, the following is a perfectly consistent position: a person can be morally blameworthy for their action, but all-things-considered, their blameworthiness ought not be communicated to them given it will lead to sub-optimal consequences. Perhaps being publicly labeled as blameworthy will make the person intractably defensive, and so engaging in a practice of blaming isn't appropriate despite the fact that the person is genuinely blameworthy. A similar position seems coherent in the epistemic case. Assuming epistemic blameworthiness is a genuine normative epistemic evaluation, one could argue that a subject's belief has the property of being epistemically blameworthy, but that it is not appropriate (all-things-considered) to engage in the social practice of holding that subject publicly to account (perhaps, again, because it will prompt a kind of undesirable irrational defensive response).

Moreover, it becomes clear that questions about the nature of blameworthiness come apart from questions about the social practice of blaming when we consider the position of some hard determinists.[77] Some hard determinists argue that subjects are never genuinely

[77] Piovarchy also draws parallels between the epistemic blame and free will/moral responsibility debates. In arguing that there is a need to investigate distinctly the nature of epistemic blame, when someone is the appropriate target of epistemic blame, and when blaming is justified, he writes: "…without a unified account it might turn out that epistemic blame is never justified. By analogy, while it is common to think of moral blame as a kind of sanction, some philosophers argue that agents are only deserving of moral blame-*qua*-sanction if their wrongdoing is an exercise of their libertarian free will. But one can believe this while also believing that, since determinism is true, no one is in fact an appropriate target of moral blame" (2021, 793).

morally blameworthy while maintaining that social practices of blaming are nevertheless appropriate or justified.[78] Roughly speaking, their position is that since hard determinism is true, people don't have free will. Here, free will is understood as voluntary control over one's actions for which once can be held morally responsible. Insofar as people aren't responsible for their actions, they aren't genuinely blameworthy for morally objectionable behaviors. Nevertheless, they could maintain that it is appropriate to engage in the social practice of blaming because it can be instrumental in encouraging morally good behavior. The social practice of blaming isn't appropriate because the persons being blamed are blameworthy. Rather, it is appropriate because blaming can be instrumental to prompting morally good behavior.[79] There is a clear epistemic analogue of this position. Epistemologists who argue against doxastic voluntarism may contend that, given we don't have voluntary control over our belief states, beliefs aren't the kind of things for which we can be held accountable. As such, they may argue that epistemic subjects are never genuinely epistemically blameworthy for holding poorly reasoned beliefs. Nevertheless, it is appropriate to engage in the social practice of epistemically blaming one another when we reason poorly because it can be instrumental in making people better reasoners (or otherwise epistemically better-off).[80]

3. Fundamental Frameworks and Epistemic Blameworthiness

With this philosophical ground somewhat cleared and organized, I turn now to my more substantive project: analyzing how epistemic blameworthiness and blaming interacts

[78] For example, see Pereboom (2014, 134).

[79] See also Pereboom (2001, 139, 147-148, 156-157); and Dennett (2015, 178-179).

[80] As I said before, in clarifying the distinction between these two projects, I don't mean to be suggesting that they are necessarily incompatible. We can imagine that an epistemologist could first give an argument for why epistemic subjects can be culpable for violating epistemic norms (perhaps by defending the claim that we have some meaningful amount of doxastic control). Then, they could offer an account of what unifies the social practice of holding one another epistemically responsible, i.e., what it takes to engage in a genuine act of epistemic blaming.

with our other epistemological commitments. In particular, I am interested in looking at how being an internalist or externalist with regards to doxastic justification informs or constrains how one should approach these two projects. I will start by considering how an epistemic internalist can and should approach questions concerning epistemic blameworthiness. Granting for the moment that there is such an evaluation, I first outline the features of epistemic blameworthiness about which I think epistemologists will likely agree. I then show that access internalists have already committed to theorizing about doxastic justification using the same normative standards that would have to govern any plausible notion of blameworthy belief. As such, I argue that access internalists must claim that, on their framework, epistemic blameworthiness just is doxastic justification. If one is an internalist, and one wants to talk about evaluations of epistemic blameworthiness, one is merely using a different term to refer to the familiar notion of doxastic justification.

3.1 A Theoretical Starting Point for Epistemic Blame

What is the general, common ground starting place for theorizing about epistemic blameworthiness? What can we all agree on? First, there seems to be general agreement that blame has to be distinct from mere negative evaluation (Boult 2021c, 518-519; Boult 2021b, 2; Hieryonymi 2004, 115-117). This is not unique to epistemologists theorizing about epistemic blame, but ethicists theorizing about moral blame as well. Many ethicists argue that a person can perform a morally wrong action, but not be morally blameworthy. For example, a utilitarian could argue that in giving to Oxfam as opposed to the Against Malaria Foundation, Matt performs the morally wrong action because the Against Malaria Foundation would have used his money more efficiently (and as such brought about greater utility). However, if we stipulate that there is no way Matt could have known about the greater efficiency of the Against Malaria Foundation, he is not blameworthy for performing an action that failed to maximize

utility. Blame, the thought goes, is more than mere "bad grading"; it has a "characteristic depth" (Rettler 2018, 2208; Hieronymi 2004, 116-117).[81] A person who is blameworthy doesn't merely violate a norm; they do so culpably. They are *responsible* for violating the norm. The norm violation is meaningfully *attributable* to them in some way.

Contemporary epistemologists who want to argue that epistemic blame is analogous though not reducible to moral blame should think that the epistemic case works similarly: for a subject to be truly epistemically blameworthy it must be the case that they have culpably violated an epistemic norm, or that they are responsible for violating the epistemic norm, or that the norm violation is attributable to them in some meaningful sense.[82] Just as debates about moral blameworthiness tend to turn on questions about what it takes to be truly culpable or responsible for acting immorally, or what it takes for an immoral action to be attributable to the agent under evaluation, so too questions about epistemic blameworthiness should turn on questions of what it takes to be truly culpable or responsible for violating an epistemic norm, or what it takes for epistemic norm violations to be attributable to the subject under evaluation.

---

[81] Piovarchy makes a similar claim in relation to epistemic blam*ing*. He argues that one can communicate a negative evaluation of another's action or belief, even an evaluation that someone has *culpably* violated a norm, without blam*ing* that person: "…neither moral blame nor epistemic blame is reducible to mere negative evaluations or disappointment. It seems perfectly coherent to say 'I'm disappointed in what he did, and he should be blamed for his wrongdoing (or bad reasoning), but I don't blame him for it'" (2021, 794). In making this assertion, I take Piovarchy to be revealing that he takes blame to be, at its core, something we do; it is not a normative evaluation that can float free of engaging in the social practice of blaming one another. It is also important to note that Piovarchy's claim is contentious. See the discussion of Fricker's view in footnote 89.

[82] This characterization of the basics of blame in general, and epistemic blame in particular, is consistent with both Brown and Boult's discussion of the topic. For example, Brown claims that "one can be blameworthy for one's beliefs in an epistemic sense, understood as being blameworthy for violating norms concerning what doxastic sates one epistemically ought to have" (2020b, 3597).

John Greco (2005) gives a similar characterization: "First, the notion of responsibility is closely tied to the notions of blame and praise. For example, judgements concerning whether a person is *morally* responsible with respect to some action or event are often equivalent to judgements about whether the person is morally blameworthy with respect to the action or event. Similarly, judgements concerning whether a person is *epistemically* responsible with respect to some belief b are often equivalent to judgements about whether the person is epistemically blameworthy with respect to b" (328).

With this theoretical common ground clearly articulated, I turn to the project of showing why the internalist must identify epistemic blameworthiness with doxastic justification. In these sections, I am going to focus on access internalism in particular. In what follows, I argue that, for the internalist, the notion of culpability, responsibility, and attributability are already embedded in their notion of doxastic justification.

3.2 The Internalist Framework and Epistemic Blame

Consider the kinds of cases that are used to motivate internalism and undermine externalism. Let's start with Stewart Cohen's New Evil Demon problem (1984). Cohen asks us to imagine two subjects, both with the exact same sensory, phenomenological experiences. Only one of these subjects' beliefs sets is reliably produced because only one subject is in the actual world; the other is being deceived by an evil demon. Cohen argues that the committed externalist reliabilist must say that, despite having identical phenomenological experiences, only the subject in the actual world has justified beliefs. Because they are not reliably produced, the beliefs held by the subject being deceived by the evil demon are not justified. Cohen finds this result implausible. He writes: "If we have every reason to believe e.g., perception is a reliable process, the mere fact that unbeknownst to us it is not reliable should not affect its justification-conferring status" (1984, 281). Here we see that a central motivation for internalism is the intuition that it is inappropriate, and unfair, to say that a subject's beliefs are unjustified when the subject could not have avoided the epistemic error. The subject being deceived by the evil demon doesn't know they are being deceived, moreover can't know they are being deceived, and are acting in a manner identical to the subject with the reliably formed beliefs. As such, the error can't be attributed to anything internal to the agent. The person being deceived by the evil demon is not at fault for the epistemic error. Though their basis for belief is corrupt, they are acting in an epistemically responsible manner in forming a belief

based on their misleading evidence. Insofar as an internalist claims that these are the features of the case that make the reliabilist's verdict untenable, they are revealing that a central internalist commitment is that judgments of doxastic justification should track evaluations of responsibility and culpability.

The central internalist intuition is that it would only be appropriate to deem a subject's belief doxastically unjustified if they are meaningfully responsible for the epistemic error that results in their possession of that beliefs. As such, the access internalist thinks it must be the case that justification is solely a matter of those internal states to which subjects have conscious access. Why? Because one can only be held responsible for that which is within one's ken. There is a clear moral analogue. It is commonly thought that ignorance vitiates moral culpability. If I didn't and couldn't reasonably be expected to know that Sally has swapped the sugar out with arsenic, I am not morally blameworthy for poisoning Paul when I, in an act of kindness, prepare him a cup of coffee. Similarly, if I didn't know and couldn't reasonably be expected to know that I am being deceived by an evil demon, then I can't be said to be doxastically unjustified when I form beliefs on the basis of my perceptual evidence.

How does this relate to our discussion of epistemic blameworthiness? Given that access internalist judgements of doxastic justification track judgements of epistemic responsibility and culpability, the internalist notion of epistemic blameworthiness can't be distinct from their notion of doxastic justification. If we can clarify the access internalist's argument for their view of doxastic justification by appealing to analogies with moral blameworthiness, then there is no room on the access internalist's framework for positing epistemic blameworthiness as a normative epistemic evaluation distinct from doxastic justification.

This point becomes clearer when we see that other access internalist arguments can be clarified by appealing to analogies with moral praiseworthiness.[83] Consider Laurence Bonjour's case of Norman the clairvoyant (1980). Bonjour asks us to imagine that Norman has a reliable clairvoyant ability of which he is ignorant. Is a true belief produced by Norman's clairvoyant abilities really justified? Bonjour argues that the committed reliabilist must implausibly respond in the affirmative. Insofar as Norman doesn't know that he has a reliable clairvoyant ability and can't produce any reasons in favor of the beliefs that the ability produces, Bonjour argues Norman is doxastically unjustified with respect to those beliefs.

Bonjour's position is similar to arguments made in defense of psychologically demanding views of moral worth or praiseworthiness (Kant 2014, 13-16; Levy 2014). Consider the titular character of Mark Twain's *The Adventures of Huckleberry Finn* (2009; first published 1884). Huck does the morally right things by helping his friend Jim escape slavery. Despite performing the morally right action, Huck falsely believes that he is doing the morally wrong thing. Some have the intuition Huck is not morally praiseworthy for helping Jim escape to freedom.[84] In defending this position, they might argue that given Huck can't offer reasons in defense of his morally right action, and moreover believes he is acting wrongly, his action is

[83] The plausibility of this section, of course, depends on the extent to which we think that praise and blame (whether of an epistemic or moral variety) are closely related. It is beyond the scope of this paper to adjudicate this issue fully. However, even if one doesn't think that praise and blame are neatly symmetrical, I contend that the claim that they are related is not deeply contentious. At minimum, questions about whether someone is praiseworthy or blameworthy are related to questions about responsibility and culpability: responsible for good or bad actions/beliefs respectively. We can accept this without further thinking that what it takes to be responsible for a good action/belief is identical to what it takes to be responsible for a bad action/belief. Accepting that praise and blame are related is, I argue, all I need to highlight that access internalists already take doxastic justification to be a matter of responsibility and culpability, and as such unable to posit epistemic blameworthiness as a distinct normative epistemic evaluation.

[84] Philosophers with this intuition likely include Kant (2014, 13-16), and those committed to Frankfurt-style views of the self, moral responsibility, and praise/blame (1971). According to Frankfurt, an action only deserves praise if a person's higher- and lower-order desires are in sync: they do the right thing, want to do the right thing, and want to want to do the right thing. Huck's lower-order desire (to aid Jim) is out of sync with his higher-order desire to do the right thing because he falsely believes that the morally right thing to do is to turn Jim in. As such, Huck wouldn't be praiseworthy on Frankfurt's view. See Arpaly and Schroeder (1999, 165-170) for discussion.

not morally praiseworthy. This, they could contend, is because moral praiseworthiness requires not only acting for the right moral reasons but knowing and being able to communicate the right moral reasons in light of which one is acting. Just like Huck can't get moral credit for doing the right thing because he cannot be said to be acting in light of the right moral reasons, so too the access internalist will want to argue Norman can't get epistemic credit for true beliefs that result from a clairvoyant ability of which he not aware. In both cases, getting credit requires that one be aware of the (good or right) reasons in light of which one is acting or believing.

With this analysis, we once again see that a fundamental internalist commitment is that doxastic justification tracks judgments of epistemic responsibility and accountability. Indeed, Bonjour explicitly says as much:

> The distinguishing characteristic of epistemic justification is thus its essential and internal relation to the cognitive goal of truth. It follows that one's cognitive endeavors are epistemically justified if and to the extent that they are aimed at this goal, which means very roughly that one accepts all and only those beliefs which one has good reason to think are true. To accept a belief in the absence of such a reason, however appealing or even mandatory such acceptance might be from some other standpoint, is to neglect the pursuit of truth; such acceptance is, one might say, epistemically irresponsible. My contention here is that the idea of avoiding such irresponsibility [that is, believing without having appropriate reasons], of being epistemically responsible in one's believings, is the core notion of epistemic justification. (1985, 8)

Given access internalist judgements of doxastic justification track judgements of epistemic responsibility and culpability, the internalist can't posit a notion of epistemic blameworthiness that is distinct from their notion of doxastic justification. At most internalists are committed to a deflationary account of epistemic blame; deflationary in the sense that any notion of epistemic blameworthiness they might discuss can be reduced to their theory of doxastic justification.

3.3 Anticipating an Objection: The Characteristic Sting of Blame

The access internalist may object to the claim that they are committed to a deflationary view of epistemic blameworthiness. They could argue that while it is true that their view of doxastic justification is similar to epistemic blameworthiness (and perhaps epistemic praiseworthiness), it is distinct insofar as only the latter (but not the former) "carries a characteristic depth, force, or sting" (Hieronymi 2004, 116-117). Consider the way in which Rettler argues that a necessary feature of (either moral or epistemic) blame is that it places a particular kind of "normative demand" on the person who is blamed (Rettler 2018, 2209). Here is her characterization of the moral case:

> "The nature of the normative demand involved in moral blame becomes clearer when we note that moral blame functions to effect a certain kind of response from the one blamed…[When I blame you,] I'm demanding a response from you: I demand that you acknowledge your failure to treat people well. Doing this requires you to recognize that you failed to act on the moral reasons you have for treating people well." (Rettler 2018, 2209)

Analogizing with the literature on moral blameworthiness, Rettler goes on to argue that "epistemic blame is a response that amounts to holding an individual responsible for failing to meet an epistemic standard" (2210). The thought is that epistemic blame, like moral blame, is a demand for redress from the subject that one believes has culpably violated a norm.

Following discussions of moral blameworthiness, access internalists could argue that while both evaluations of doxastic justification and blameworthiness track epistemic responsibility and culpability, only the latter places a "normative demand" on the evaluated subject (Rettler 2018, 2209). The thought is something like the following: When one is blameworthy, one isn't merely graded badly because one has failed to meet some normative standard. Rather, one is graded badly and faces a demand for explanation, or an expectation

that one acknowledge one's failure, or a call to promise that one not act similarly in the future.[85] I understand the idea to be that blame is fundamentally a social sanction that necessarily carries a call to action (apology, changes to future behavior, etc). One's belief, the access internalist could argue, may be doxastically unjustified but it is only epistemically blameworthy when it appropriate to level a social sanction that places a normative demand on the evaluated subject.

I have several concerns with this access internalist strategy for resisting the previous section's argument. The first questions what could motivate attaching the social sanction/normative demand element to epistemic blameworthiness, but not the access internalist notion of doxastic justification. Presumably, the reason that philosophers want to argue that blame "carries a characteristic depth, force, or sting" is that one is only blameworthy when one is culpable for a norm violation (Hieronymi 2004, 116-117). One deserves social sanction and faces social demands for appropriate redress when one is blameworthy because, in such cases, one is responsible, or on the hook for one's norm violation. It is the culpability that makes the sting of blaming appropriate. Given that the access internalist thinks that doxastic justification is a matter of responsible or culpable norm violation, it is not clear what, if anything, could justifying attaching a social sanction/normative demand to a notion of epistemic blameworthiness but not their theory of doxastic justification.

My second concern with this strategy is that it conflates the two inquiries disentangled at the start of this paper: the inquiries into blameworthiness and blaming. On one reading, the most that can be done with the social sanction-normative demand strategy outlined above is the following: On the access internalist framework, evaluations of epistemic blameworthiness reduce to evaluations of doxastic justification. However, when one asserts that a subject's

[85] This sentiment is echoed in the moral literature. See Hieronymi (2014, 117); McKenna (1998); Strawson (1974, 6-7, 14-16).

belief is doxastically unjustified, one engages in a practice of epistemic blaming insofar as one places a social sanction on the subject and demands redress for their culpable epistemic error. This is fine so far as it goes, but it doesn't undermine my argument that any internalist notion of epistemic blameworthiness is ultimately reducible to discussions of doxastic justification. Perhaps what my argument has done is clarified that, if an internalist is interested in theorizing about epistemic blame, the novel issue with which they can engage concerns the nature of a social practice, not the necessary and sufficient conditions of special normative epistemic evaluation.[86]

4. How Epistemically Illuminating is a Conceptual Analysis of Epistemic Blaming?

This would be an interesting result. After all, contemporary epistemologists debate what constitutes a genuine instance of epistemic blaming. In recent literature, two prominent views have emerged (Brown, 2020a; Boult, 2021a). Each takes an account of moral blaming and amends it to provide an account of epistemic blaming.

Brown starts with George Sher's account of moral blame: the position that a genuine act of blaming involves having a particular belief-desire pair that explains the dispositions that we typically have when we morally blame other people. According to Sher's account of the moral case, the relevant belief-desire pair is "the belief that someone has acted badly or has a bad character, and the desire that they not have acted badly or not have had a bad moral character" (Brown 2020a, 395; Sher 2006, 103). Believing that someone has acted badly and having the desire that they hadn't explains why we tend to reproach those we blame, why we think they ought to be punished in some cases, and why we might ask for an apology. Brown amends this to craft an account of epistemic blaming. She argues that there is an epistemic

[86] And I think it is worth noting that this is not an unhelpful intermediary conclusion, given contemporary epistemologists do take themselves (at least at some instances) to be pursuing the latter project. See Brown (2020a) and Boult (2021a).

analogue of the relevant belief-desire pair that concerns "believing badly" (Brown 2020a, 399). When we epistemically blame others, we believe that the person being blamed has believed badly, meaning they have beliefs that "violate epistemic norms without excuse", and we desire that they hadn't done so (399). The frustrated desire explains the dispositions, i.e. negative emotions and related behavior, Brown claims are typically associated with epistemic blame.

In contrast, Boult modifies T. M. Scanlon's notion of moral blame into an account of epistemic blame (Boult 2021a; Scanlon 2008). Scanlon argues that understanding blame and blaming requires focusing on relationships. In an epistemically analogous vein, Boult argues that epistemic blaming occurs when one modifies the "intentions, expectations, and attitudes you have towards [the blamed] person, in a way made fitting by the judgement that they are blameworthy" (Boult 2021a, 11). Importantly, Boult's view makes room for the kinds of reactive attitudes that are often taken to be characteristic of moral blame, without making them central as Brown's account does.

I would like to offer some reasons for thinking that normative epistemologists should set conceptual analyses of epistemic blaming to the side. I want to argue that we should not expect conceptual analysis of epistemic blaming to be deeply philosophically illuminating. Moreover, I want to argue that investigations into what genuine epistemic blaming looks like is better understood as within the purview of psychologists and anthropologists, not armchair epistemologists.

My argument starts by questioning why we should think that there is something that binds together practices of epistemic blaming (or blaming generally) other than the fact that they are explained by a judgement that someone is blameworthy. Why do we need to adjudicate between whether modifying a relationship, or having a particular reactive attitude, is a necessary constitutive feature of blaming? So long as the reaction in question is explained by

the judgement that a person has acted or believed in a blameworthy fashion, is it not an instance of blaming? Blaming responses and practices are incredibly diverse. It seems perfectly reasonable that a person's blaming response depends on facts about their particular dispositions: how quick they are to form and communicate reactive attitudes like indignation and resentment; how they view their relationships and associated obligations of loyalty and deference. If we accept that there is a diverse set of possible responses to these questions, then we should also accept that blaming responses can vary dramatically, despite all counting as blaming responses.[87, 88]

Consider a person who is not disposed to form and communicate reactive attitudes, and not disposed to alter their relationships with people who they believe have culpably violated norms. We might think it unusual that they don't have strong reactions to culpable wrongs (particularly when they themselves are the harmed party). We might think this person acts irrationally when they don't reduce their trust in the person who has culpably violated an epistemic norm. However, I contend there is no reason we shouldn't understand such a person as blaming when they dispassionately and without consequences (i.e., without modifying the

[87] Though Piovarchy's project (2020) is to offer a novel account of epistemic blame (an agency-cultivation account), he seems to agree with this point. Consider what he says when explaining how his position answers the conceptual question regarding what epistemic blame is. He writes: "epistemic blame consists in a cognitive judgement that someone is blameworthy—namely, that they are an appropriate target of a certain class of interpersonal reactions for violating epistemic standards—which in turn creates a disposition to engage in blaming reactions, the expression of which communicates to the agents our dissatisfaction with their violation of those standards. This is particularly appealing because it explains why epistemic blame manifests in a variety of forms, and it overlaps with our conception of moral blame" (2020, 12). Like Piovarchy, I think what unifies blaming responses is simply the judgment that someone is blameworthy. I agree that the blaming reactions that come along with those judgements can be varied. I am adding that to give a unified account of them from the philosopher's armchair would be to venture into psychology, sociology, and anthropology without using the proper research methodology.

[88] I am not alone in thinking that we cannot subject our practices of blam*ing* to rigorous conceptual analysis and come up with a philosophically illuminating view. Fricker similarly argues that "we should not expect any such analysis to be very illuminating, owing to the fact that blame is significantly disunified, and is therefore likely to have distinctive or otherwise central features that may not be present in all instances" (2016, 166). She instead uses a "paradigm based method" (166). She does endorse a paradigm case approach to theorizing about the point or value of blame. By examining "Communicative Blame" she thinks we can arrive at an account of blame that "aims to explain the nature of the practice in all its internal diversity" (166).

relevant relationship) communicate that they think someone has believed badly, or acted wrongly.[89]

Diverse blaming responses are unified because they are explained by the fact that, in each case, an individual has formed a belief that someone else is blameworthy.[90] But this is not a particularly interesting or philosophically substantive analysis. If we are interested in the myriad of ways people act when they formed such a beliefs, I think we should turn to psychologists and anthropologists.[91]

5. Why does this matter?

At this juncture, some may wonder why this all matters. Why care that the access internalist can't make a meaningful distinction between epistemic blame and doxastic justification? First, it is important because the contemporary literature on epistemic blame is, I would argue, in large part divorced from the existing literature on doxastic justification. We might think it is worth it to save the access internalist the trouble of re-theorizing the same normative evaluation over again. But in addition to these pragmatic concerns, I will argue that there are other more philosophically substantive benefits to recognizing that the access internalist's notion of doxastic justification just is a notion of epistemic blameworthiness. It prompts us to consider the following: Given the access internalist's notion of doxastic justification just is a notion of epistemic blame/praiseworthiness, can proponents of other views of doxastic justification, in particular externalists, really have a notion of epistemic blame

[89] Fricker would agree. She writes, "[s]omtimes blame is little more than a dispassionate judgement that someone is blameworthy, the merest answer to the question 'Whose fault is it?'" (2016, 167).

[90] Fricker would also agree with this point. She writes that a "minimal definition of blame" is likely "*a finding fault with someone for their (inward or outward) conduct*" (170).

[91] I am concerned that when we start to give a more specific account of what interpersonal behavior counts as "genuine" blaming, we risk allowing cultural imperialism to sneak into our ethical and epistemological theorizing. Different cultures may have different views about what reactive attitudes are (in)appropriate in different relationships. It may be the case that which behaviors are explained by the judgement that someone is blameworthy are culturally diverse.

while rejecting access internalism as a theory of doxastic justification? Consider the reliabilist externalist who wants to adopt a pluralistic epistemic framework on which doxastic justification is an externalist reliabilist notion, and epistemic blameworthiness has a more internalist flavor. Will the arguments that this externalist reliabilist offers against adopting access internalism as a theory of doxastic justification similarly undermine the plausibility of epistemic blameworthiness as a genuine normative epistemic evaluation? In what follows, I investigate whether externalists can have this pluralistic epistemic framework on which doxastic justification and epistemic blameworthiness are two distinct normative epistemic evaluations.

6. Can Externalists Have their Cake and Eat it Too?

Accepting this kind of pluralistic epistemic framework should be attractive to many externalists because it can ostensibly be used to explain internalist intuitions while preserving an externalist theory of doxastic justification. The central kind of case motivating this view is one in which the evaluation of blameless but unjustified belief is appropriate – i.e., cases in which a subject's belief is unreliably formed, though the unreliability is not the result of the subject's epistemically irresponsible behavior. For example, an externalist defender of a pluralistic epistemic framework would say that the brain in a vat has unjustified but epistemically blameless false beliefs about the external world, as does the subject being deceived by Cohen's evil demon. The subjects in these cases are not failing to consider relevant evidence, they are not willfully naïve, and they are not problematically cavalier about the truth. Rather, the unreliability of their belief-forming process is explained by events and states of affairs for which they cannot be held responsible. Thus, while they have doxastically unjustified beliefs, they are epistemically blameless with regards to those beliefs. This is an

attractive, intuitive analysis for many externalists. It seems like the way they can have their proverbial cake and eat it too.[92]

The motivation an externalist has for adopting a pluralist framework seems akin to that an actual results utilitarian has for positing a notion of epistemic blame that tracks agents' intentions and character.[93] The actual results utilitarian is committed to saying that the person who didn't maximize consequences acted wrongly. However, she can claim that this wrong action might be blameless insofar as the person was faultlessly ignorant of the consequences of her action. In crafting a theory of moral blame, the actual results consequentialist can claim that her view doesn't give implausible verdicts. The thought is that actual result utilitarianism isn't too demanding; we just have to remember it is a theory of right action, not praise or blame.

6.1 What Evaluative Pairings are Possible?

If externalists are going to posit epistemic blameworthiness as a genuine normative evaluation distinct from doxastic justification, surely, we should expect there to be other types of cases in which the explanation given by this epistemic framework is appropriate and

---

[92] It should be noted that many philosophers have tried to construct epistemic frameworks that combine the insights of externalism and internalism. See for example, Ernest Sosa's distinction between two different grades of knowledge, "animal" and "reflective" (2007). Juan Comesaña defends a view that combines elements of externalist reliabilism and internalist evidentialism (2010). See also Goldman 1993, Williams 1977, and Kornblith 1985 (though he no longer holds this view).

[93] Indeed, John Stuart Mill makes a distinction between the standards of right action, and what an agent's motivation tells us about an action or an agent. He writes:

> They say it is exacting too much to require that people shall always act from the inducement of promoting the general interests of society. But this is to mistake the very meaning of a standard of morals, and confound the rule of action with the motive of it. It is the business of ethics to tell us what are our duties, or by what test we may know them; but no system of ethics requires that the sole motive of all we do shall be a feeling of duty; on the contrary, ninety-nine hundredths of all our actions are done from other motives, and rightly so done, if the rule of duty does not condemn them. It is the more unjust to utilitarianism that this particular misapprehension should be made a ground of objection to it, inasmuch as utilitarian moralists have gone beyond almost all others in affirming that the motive has nothing to do with the morality of the action, *though much with the worth of the agent.* (29; emphasis mine)

Thank you to Pete Graham for pointing this out to me. See Arpaly (2002, 224-224) for discussion as well.

illuminating. In other words, other combinations of the two epistemic evaluations, doxastic justification and epistemic blameworthiness, should be theoretically plausible and have explanatory value for the externalist. After all, blamelessness seems an empty evaluation if one couldn't, in principle, have been blameworthy.

The following table is intended to help us visualize this step of the project.

	Doxastically Justified (i.e., formed by a reliable process)	Doxastically Unjustified (i.e., formed by an unreliable process)
Epistemically Blameworthy	?	?
Epistemically Blameless	X	Person being deceived by Cohen's new evil demon, brains in vats, etc.

Figure 1.

In the next section, my focus will be examining the plausibility of the epistemically blameless unjustified combination. Before doing so, I want to briefly explain why I am choosing not to treat the combination of justified but epistemically blameless as a live option. Consideration of a moral analogue is helpful. Recall the discussion of actual results utilitarianism above.[94] Imagine a person who successfully maximizes utility – a person who acts rightly by the lights of the utilitarian. What does it mean to say such a person is blameless? To be blameless is to be off the hook. But in what sense could one have been on the hook for a morally right action? We might be tempted to say that blameless right action is an instance of praiseworthy right action, but this would be too hasty. It seems likely that praiseworthiness requires more than mere blamelessness. All to say, I think there are good reasons for setting this combination to the side.

[94] Hereafter, I will use "utilitarianism" to refer to actual results utilitarianism.

6.2 Unjustified and Epistemically Blameworthy

Stipulate for the moment that we are externalist process reliabilists. Let's start by considering how comfortable we are with cases in which a subject has an unjustified belief, and they are epistemically blameworthy for holding that belief. On its face, this case seems relatively straightforward: a situation in which someone forms a belief in an unreliable fashion, and the unreliability of the belief-forming and/or sustaining process is explained by their own epistemically irresponsible behavior, i.e., epistemically bad behavior for which they are at fault. The following caricature of a dogmatic climate change denier might describe a subject deserving of this evaluation:

> **Dogmatic Climate Change Denier:** Megan firmly believes that if there is any climate change, it isn't caused by human fossil fuel consumption. She believes that any changes to the planet's climate are completely attributable to the planet's natural cooling and warming cycles. Although she looked at relevant information and data in forming this belief, she lacks the relevant expertise to reliably form true beliefs on the basis of such data. As such, her belief is unjustified. Moreover, one way Megan continues to sustain her false unjustified belief is by consciously refusing to consider evidence contrary to her views. For example, she knows of articles in genuinely reputable journals that she herself trusts. She knows that reading these articles would likely undermine her views. However, in full awareness of what she is doing, she refuses to engage with these articles.

Insofar as Megan is consciously refusing to engage with good evidence that would undermine her view, it seems appropriate to say that she is acting in an epistemically irresponsible fashion; her epistemic behavior is not only bad, she is also on the hook for it. As such, the evaluation of unjustified and epistemically blameworthy might seem fitting. Are we are getting any more explanatory milage by using two normative epistemic evaluations instead of one? The defender of the externalist pluralistic epistemic framework I outline above wants to answer in the affirmative: one evaluation tracks reliability (doxastic justification), the other epistemic responsibility (epistemic blameworthiness). Moreover, this is a meaningful thing to say, given

judgements of (un)reliability and (ir)responsibility for (un)reliably can come apart (as in the case of the brain in a vat).

	Doxastically Justified (i.e., formed by a reliable process)	Doxastically Unjustified (i.e., formed by an unreliable process)
Epistemically Blameworthy	?	Dogmatic climate change denier.
Epistemically Blameless	X	Person being deceived by Cohen's new evil demon, brains in vats, etc.

Figure 2.

6.3 The Possibility of an Externalist Norm of Epistemic Responsibility that Tracks Internalist Intuitions

So far, so good. However, if this is going to work, we need to investigate whether the externalists can defend a norm of epistemic responsibility that not only tracks internalist intuitions but also is compatible with their other, externalist commitments. In this section, I will argue that if an epistemologist's commitment to externalism involves what Amia Srinivasan, in her defense of "Radical Externalism", terms "Anti-Cartesianism", then it is not obvious that an externalist can consistently endorse a norm of epistemic responsibility (which is needed to make the pluralistic epistemic framework work) (2020, 274).

Anti-Cartesianism is the view that "[t]here are no transparent conditions", meaning that "for any condition we can sometimes fail to be in a position to know whether we are in it" (Srinivasan 2020, 274). This is uncontroversial with respect to conditions or states of the external world. I could fail to be in a position to know that it is raining even if it in fact is. This could be because it started raining when I entered the cinema and the movie I am watching hasn't yet finished. Anti-Cartesianism is controversial, however, because it amounts to the claim that we can fail to know our own psychological and phenomenological states. If Anti-

Cartesianism is true, I may not know what I myself believe about the aesthetic merits of the movie I am watching.

Anti-Cartesianism has untenable consequences for many theorists, access internalists in particular. As Srinivasan notes, for a particular kind of epistemologist, if it is the case that Anti-Cartesianism is true, the very notion of genuine epistemic norms starts to fall apart. On the access internalist's view, genuine norms must concern transparent states. If a genuine norm doesn't concern transparent states, then we can't make a principled distinction between subjects who accidentally conform with a norm, and those that follow, or act in light of the norm. Their argument is that one can only be said to follow an epistemic norm by forming a belief if one has access to their basis for belief; one can't get credit for following a norm by forming a particular belief if is one isn't aware of the why one holds the belief. As Srinivasan puts it, for access internalists "only mental states that possess a crucial property, namely transparency, and only norms that feature transparent states can meet some basic desiderata of norms" (2020, 276). This is because Anti-Cartesianism entails "Anti-Lucidity", the view that "there is no norm such that a competent agent who knows the norm is [automatically or necessarily] in a position to know of every basic action available to her whether it would in conformity with that norm" (277).

Recall that we are exploring whether an externalist can coherently say that the dogmatic climate change denier has a belief that is not only unjustified, but for which he is epistemic blameworthy. Recall that the motivation this kind of pluralistic epistemic framework is the ability to preserve an externalist notion of justification while explaining internalist intuitions. This is accomplished by having a theory of epistemic blameworthiness that tracks an internalist-friendly norm of epistemic responsibility. Importantly, an internalist-friendly norm of epistemic responsibility, a norm that can really capture and explain internalist

intuitions, must be a lucid norm. A lucid norm is such that it must be the case that "a competent agent who knows the norm is in a position to know of every basic action available to her whether it would be in conformity with the norm" (Srinivasan 2020, 277). But here arises the problem for the externalist looking to craft this pluralist epistemic framework: insofar as they are committed to Anti-Cartesianism, and by extension Anti-Lucidity, they can't consistently claim that there is a norm of epistemic responsibility that is always lucid. In light of this, we might question whether an externalist can coherently endorse a normative evaluation of epistemic blameworthiness that tracks internalist intuitions.

In some ways, this result is surprising given one of Srinivasan's projects is arguing that Anti-Cartesianism, and by association Anti-Lucidity, entails "Blameless Violability": the claim that "[a]ny norm can be blamelessly violated by a competent agent who knows the norm" (Srinivasan 2020, 285; emphasis mine). She argues that Blameless Violability follows from Anti-Cartesianism insofar as Anti-Lucidity means that even a competent agent who knows a given epistemic norm can faultlessly violate that norm because no conditions – neither their psychological states nor states of the external world – are necessarily transparent. Given that bad luck could result in the opacity of a relevant condition, it is always a possibility that an agent can faultlessly violate an epistemic norm that they know, and as such be epistemically blameless with respect to the relevant belief.

But our focus here is on the converse of epistemic blamelessness: epistemic blameworthiness. I have argued that the notion of epistemic blameworthiness the externalist wants requires an internalist-friendly lucid epistemic norm of responsibility, and that this is potentially at odds with their commitment to Anti-Cartesianism. This suggests there isn't a notion epistemic blameworthiness that the externalist can consistently endorse that satisfies the motivation for adopting the pluralistic framework. And as I mentioned at the outset of

this investigation, this has an impact on the notion of "Blameless Violability" and the pluralistic framework in general. After all, if it is not in theory possible for a subject to be at fault for violating an epistemic norm, what does it mean to say that that subject faultlessly violated an epistemic norm? An evaluation of epistemic faultlessness, or blamelessness, is only illuminating and explanatory if there is a corresponding evaluation of epistemic fault or blame. What does it mean to be epistemically excused if one couldn't, in principle, have been on the epistemic hook?

6.4 Anticipating an Externalist Reply: Contextual Transparency

A defense of the pluralistic epistemic framework may still be available. Anti-Cartesianism and Anti-Luminosity are consistent with both conditions being "contextually transparent" and norms sometimes (even often) being lucid (Srinivasan 2020, 278). Yes, it is always a possibility that we fail to know that we are in a particular state or condition. Consequently, we fail to know whether our beliefs conform to epistemic norms. Despite this, states and conditions can be transparent in certain contexts; similarly, norms can be lucid in certain contexts. The externalist aiming to preserve epistemic blame as a genuine normative epistemic evaluation could argue that the norm of epistemic responsibility only applies when we know our belief and the basis for our belief (or ought to have known our belief and the basis for our belief). In such cases, we know (or ought to know) whether our epistemic behavior conforms to the relevant epistemic norms. In other words, we can only be epistemically blameworthy when we are in contextually transparent conditions. This, the externalist could argue, is what is going on in the case of the dogmatic climate change denier.

Nonetheless, the externalist's approach won't be able to accommodate all the internalist intuitions the externalist might like to explain. This is an issue, given that the motivation for developing this pluralistic epistemic framework is accounting for internalist

intuitions while preserving an externalist theory of doxastic justification. Consider the traditional internalist analysis of Bonjour's Norman the clairvoyant case. The internalist wants to say that Norman is doxastically unjustified in believing that the president is in New York because the belief is the result of epistemically irresponsible behavior, despite being reliably produced. What analysis should the externalist committed to the pluralistic epistemic framework offer? Rather than puzzling out how to resist Bonjour's claim—that process reliabilism is committed to saying that Norman's clairvoyant beliefs are justified—externalists could instead grant this verdict. Doxastic justification tracks reliability while epistemic blame tracks epistemic responsibility. Insofar as Norman's belief is reliably produced, externalists should say he is doxastically justified. Insofar as they are trying to explain internalist intuitions about epistemic responsibility, they should say he is epistemically blameworthy.

	Doxastically Justified (i.e., formed by a reliable process)	Doxastically Unjustified (i.e., formed by an unreliable process)
Epistemically Blameworthy	Maybe Norman the clairvoyant?	Dogmatic climate change denier.
Epistemically Blameless	X	Person being deceived by Cohen's new evil demon, brains in vats, etc.

Figure 3.

We must pause here and think about what the externalist can say to explain why Norman is epistemically blameworthy with respect to his clairvoyant beliefs. Let's consider the contextual transparency strategy discussion above. On this view, the thought would be that Norman's belief that the president is in New York is epistemically irresponsible because this is a case with contextually transparent conditions. As such, the relevant epistemic norm is lucid. The conditions are such that Norman ought to have known he violated an epistemic norm.

But can the externalist really say that Norman ought to have known that he is believing in an epistemically bad way? The externalist can't say that Norman ought to have known his

clairvoyant abilities were unreliable, because his clairvoyant abilities are reliable. Moreover, it doesn't seem the externalist can say that Norman ought to have believed his clairvoyant abilities are unreliable. If Norman ought to have had this belief, it is presumably because he had (or ought to have had) a defeater: evidence that his clairvoyant abilities are unreliable. The notion of a defeater can only do explanatory work for the process reliabilist here if it is the case that believing in the face of a defeater is epistemically bad – that is, believing in the face of a defeater is epistemically bad because it is unreliable. However, if this explains why Norman ought to have believed that his clairvoyant belief was epistemically bad, then by the process reliabilist's lights Norman's belief should be considered unjustified.

I suspect this doesn't just mean that Norman is an unsuitable candidate for the dual evaluation of justified but epistemically blameworthy. Rather, it means that the process reliabilist can't plausibly evaluate any subject's belief as justified but epistemically blameworthy. To do so, a process reliabilist would have to be able to say both (1) that a subject's belief that *p* was reliably formed and (2) that a subject is meaningfully criticizable because they ought not have believed their belief that *p* was reliably formed. However, any plausible argument a process reliabilist will give for the claim that a subject should not believe that their belief that p was reliably formed will amount to an argument for thinking that the belief that p is doxastically unjustified by the process reliabilist's own lights.

To be clear: it does not follow from this conclusion that process reliabilists must deny the existence of misleading higher-order evidence concerning belief-forming process reliability. Rather, it amounts to saying that if a reliabilist wants to claim that we can criticize a subject for ignoring such evidence, then they are committed to saying that ignoring such evidence renders one's belief-forming process unreliable and as such that process is unable to confer positive justification on output beliefs.

	Doxastically Justified (i.e., formed by a reliable process)	Doxastically Unjustified (i.e., formed by an unreliable process)
Epistemically Blameworthy	X	Dogmatic climate change denier.
Epistemically Blameless	X	Person being deceived by Cohen's new evil demon.

Figure 4.

It is not obvious that this result is damning for the externalist's defense of the pluralistic epistemic framework. One could argue that it need not be the case that every combination of evaluations be plausible for the pluralistic framework itself to be plausible. After all, consider the ethicist who posits that evaluations of moral rightness, wrongness, and permissibility are distinct from evaluations of moral blameworthiness and praiseworthiness. Even they might say that there is something odd about judgments of morally right and blameless action, and blameworthy right action.[95] That said, it is worth noting that the notion of epistemic blame can't do all the work that the externalists want it to do.

But perhaps this should prompt us to consider whether Bonjour's Norman the clairvoyant is the right kind of case to test this pairing of epistemically blameworthy but doxastically justified. Let's think about a case of morally blameworthy right action. Consider: a person who consciously intends to harm another person but fails to do so because they were mistaken about the consequences of their actions. According to an actual results consequentialist, such a person may end up performing the morally right action but nevertheless be blameworthy in some sense. Imagine that Becky knows that Sarah has a peanut allergy, and Becky wants to harm Sarah by giving her a cookie with peanuts in it. Becky buys the peanut-laced cookies and leaves them in the cupboard. Unbeknownst to Becky, another roommate comes along and eats the peanut cookies and being a conscientious roommate,

[95] Though I think such cases of morally blameworthy right action are plausible. See the cookie case below.

replaces the cookies. However, the new package of cookies happens to be peanut-free. Becky offers one of the cookies to Sarah. Sarah has no allergic reaction to the peanut-free cookie, and in fact, the gesture (that Sarah perceives as kind) really makes her day. Sarah very much enjoys the cookie and feels great because she believes that someone went out of their way to give her a special treat. A committed, actual results consequentialist might argue that, despite the malicious intent, Becky's action was morally permissible. Nevertheless, some negative evaluation is appropriate given Becky's malicious intent, and blameworthy seems to do the trick. The externalist defending the pluralistic epistemic framework might think we need a case analogous to this to test the plausibility of the epistemically blameworthy/doxastically justified pairing.

But what would such a case look like – perhaps a case where someone happens to form a belief using a reliable process, even though they were actively trying to avoid the truth? This case seems hard to imagine. To clarify: I don't mean to say that it is unusual for epistemic subjects to actively avoid the truth, or desire to believe a falsehood.[96] It is just difficult to imagine a case in which one fails to achieve intentional ignorance in a way that warrants the label of epistemically blameworthy but doxastically justified.

Consider what cases of failed intentional ignorance would look like. Someone might try to achieve intentional ignorance by avoiding evidence that would prompt the formation of the undesired belief. We can imagine that they are forced to confront the evidence by a persistent interlocutor, after which they can't help but form the undesired belief. Their reliable perceptual and cognitive beliefs entail that the undesired belief is the output of a reliable belief-forming process, and thus it is doxastically justified. While it certainly seems appropriate to

[96] This seems to be what Michelle Moody-Adams has in mind when she discusses "affected ignorance": "choosing not to know what one can and should know" (1994, 297).

say that they are not epistemically praiseworthy with respect to the belief, it is less clear whether it is appropriate to say that they are epistemically blameworthy with respect to the belief. If anything, it maybe feels appropriate to say something about this person's overall epistemic character. Moreover, I think it will be difficult to come up with a case in which our intuitions are distinctly epistemic, and not influenced by the moral stakes of the case. Even if we could come up with a case, we would have to consider: if the epistemic behavior associated with actively avoiding the truth is bad enough to warrant inclusion in our epistemic evaluation, why wouldn't it render the belief-forming process unreliable?

Alternatively, someone might want to argue that I am making a mistake by focusing on cases of moral and epistemic blameworthiness where there is conscious intent. After all, someone could arguably be morally blameworthy not for malicious conscious intent, but for a failure to demonstrate the appropriate moral concern. Here, we can think of a long-term committed romantic partner who doesn't consciously and maliciously intend to sabotage their partner's birthday, but rather forgets (Sher 2009). Perhaps we need to consider a case of right action with this kind of lack-of-proper-attention blameworthiness. This seems hard to come by. At least by my lights, lucking into right action while having a criticizable lack of awareness doesn't prompt a clear intuition of moral blame.

Imagine a long-term committed romantic couple: Ellie and Leigh. Ellie forgets that today is Leigh's birthday. However, Ellie decides to surprise Leigh with a present just because she wants to do something nice for her partner. The present is well-received; it makes Leigh happy. I can certainly understand why we might think this action isn't (at least exceedingly) praiseworthy. Inasmuch as forgetting is an action that can be ethically evaluated (and this seems controversial), we could say that Ellie's forgetting is both morally wrong and blameworthy. Insofar as forgetting leads to Ellie doing, or failing to do, something that makes

Leigh sad, Ellie seems to be doing something morally wrong and blameworthy. But is giving the "something nice" present a morally blameworthy right action? I don't think the charge is appropriate.

This analysis reveals problems with the analogies drawn by a proponent of the externalist pluralistic epistemic framework – or at least, problems with the analogies drawn between consequentialist ethics and externalist epistemology. By the process reliabilist's lights, doxastically justified belief isn't neatly analogous to the actual result consequentialist's notion of right action. When you are an actual results consequentialist, you think that it is possible to luck into performing the right action, as in cases where your action produces optimal, although unintended or unintentional, consequences. It is precisely because you can luck into performing the right action that notions of moral praiseworthiness and blameworthiness are useful. A subject who lucks into the right action may not be praiseworthy for performing the morally right action. But by the process reliabilist's lights, you can't luck into a doxastically justified belief in the same sense. Indeed, one way of understanding the project of theorizing about justification is as theorizing about the anti-luck condition on knowledge. At its heart, the process reliabilist's notion of doxastic justification is etiological, while the actual results consequentialist's notion of right and wrong action is, well, consequentialist; the former is partially etiological, the latter is purely forward-looking.

Does this disanalogy pose a problem for the externalist looking to defend a pluralistic epistemic framework? I think it does. There are roughly two intimately related components to understanding the connection between acting for the right reasons and responsibility: (non-)accidentality and expression of character.[97] As to (non)accidentality: when we want to evaluate

[97] This way of carving of the issue is drawn from Arpaly (2002), and helpful discussions with John Robison.

whether someone is blameworthy for performing the wrong action, we want to know whether it was just chance that they performed the wrong action.[98] Turning to the second component, expression of character: here, the relevant concern is about whether the action says something about the person who performed it.[99, 100]

If we focus on the first component, (non-)accidentality, and think that this is centrally what is interesting about moral responsibility then here is my contention: we should straightforwardly think that moral responsibility, and accordingly praiseworthiness and blameworthiness, are analogous to epistemic justification regardless of whether we are internalists or externalists.[101] If this is the case, then just as I argued with the access internalist above, the externalist can't have a notion of epistemic blameworthiness, and for the same reason: any notion of epistemic blame they can endorse will collapse into their notion of epistemic justification.

---

[98] If someone performs the wrong action for malicious reasons, then it is not just chance that they performed the wrong action. If someone performs the wrong action but their intentions were beneficent, then it seems like an accident that they performed the wrong action.

[99] Does it reveal or reflect something about their character? If it does, then they are responsible, and accordingly can be praised or blamed. If not, then the responsibility judgment seems inappropriate, as do praising and blaming attitudes.

[100] Of course, (non-)accidentality and expression of character are deeply connected. If it was merely an accident that one performed the wrong action, then the wrong action doesn't seem to reflect anything about someone's character. However, I think they are distinct. After all, it seems like someone can perform an action non-accidentally, on purpose, even if the action isn't meaningfully representative of who a person is. I am thinking of the kind, soft-spoken person who, in moment of intense and unusual frustration, purposefully says something that is less than kind. In this sort of case, which I believe is not uncommon, we may hold the person responsible, and blame them, though we would perhaps limit the extent to which we affectively express those judgments. Of course, someone might say that there is a sense in which the less-than-kind utterance was "just an accident" given it is out of character. However, in a straightforward sense, it isn't. Surely an action can be intentional, and therefore non-accidental, without being deeply expressive of one's personality and character. Indeed, Arpaly similarly notes that we can separate out considerations of (non-accidentality) from consideration of character or depth of concern: "A traditional objection states that virtue ethics is committed to excusing his action as "out of character," but if "in character" does not mean "predictable" or "in keeping with historical trends," then "unpredictable" and "out of keeping with trends" does not always mean "out of character," either" (2002, 96).

[101] One might say: Even if we grant that moral responsibility judgments track epistemic justification, why would we think that puts any pressure on the plausibility of a pluralistic epistemic framework. After all, we have notions of moral responsibility *and* moral blame/praise. Doesn't this show the responsibility is distinct from blame/praise? Though we use different terms, I would contend that there isn't a deep distinction here. I think moral praise and blame track responsibility for good and bad respectively. It is not that they are distinct from responsibility, it is rather that when we are talking about responsibility, we are talking about responsibility *for something*, and that something has a particular moral valence.

To see why this is the case, we must look at accounts of moral responsibility, praise, and blame, that are sympathetic to the psychological considerations that are core to the motivations of many externalists. These are accounts of moral responsibility that take seriously the idea that our own psychological states, our reasons for acting, aren't transparent to us. Consider Nomy Arpaly's analysis of moral worth, and the associated notion of moral responsibility, praise, and blame (2002).[102] On her view, whether an action has moral worth depends in part on whether the agent was responsive to the right moral reasons. Importantly, this reasons responsivity needn't be conscious, or something of which the agent is aware.[103] I argue that this kind of moral reasons responsivity is akin to a kind of moral reliabilism.[104] The general idea is that moral worth, praise, and blame are determined by whether the morally right action is the product of moral mechanisms that reliably produce good or bad actions, and that are rooted in concern for morality or proper values. If one accepts this view of moral responsibility, praise, and blame, then it is unsurprising that an externalist should think that moral responsibility is analogous to epistemic justification (and correspondingly, that there can't be a meaningfully distinct notion of epistemic blame). [105]

---

[102] Indeed, Arpaly aligns herself with externalism. As will be discussed, her view is that moral worth is a question of acting in responsive to the moral reasons out of moral concern. Importantly, on her view responsivity to the moral reasons is not psychologically demanding. In defense of this view, she writes: "The idea that we can sometimes act for moral reasons without knowing that we act for moral reasons is not strange when posed against the background of epistemology and psychology, where many have maintained that we can know without knowing that we know, believe without believing that we believe, or act for a reason without knowing that we act for a reason" (2002, 230).

[103] Arpaly goes so far as to say that the reasons responsiveness that underlies praise and blame doesn't have to involve even subconscious knowledge of the right or wrong reasons. In her analysis of the case of Huckleberry Finn, she writes: "Contra Rosalind Hursthouse, my point is not simply that Huckleberry does not have the belief that his action is moral on his mind while he acts. He does not have the belief that what he does is right *anywhere* in his head—this moral insight is exactly what eludes him…On this reading, he is not a bad boy who has accidentally done something good, but a good boy with imperfect knowledge" (2002, 229-230).

[104] See also Robison 2020.

[105] Importantly, in defending a non-psychologically demanding view of moral worth, Arpaly (and those with similar views) need not say that being consciously aware of the reasons for which one is acting is always irrelevant to determinations of responsibility, praise, and blame. Arpaly contends that blame is a degreed rather than categorical notion. The depth of one's moral concern can influence the degree to which one is praise or blameworthy. Sometimes, though not always, conscious awareness of one's reasons for action are an indication of the level of moral concern one has. For example, it is not that the person who isn't responsive to the moral

Let's turn our attention to the other component of moral responsibility, expression of character. Some will likely want to argue that when we attend to this component, we will not get the same result.[106] I don't think this is the case. Questions about expressions of character are in part etiological, just like evaluations of reliability. When we are evaluating the reliability of a belief-forming process, we are the evaluating subject's own cognitive faculties; we are asking whether something stable and truth-conducive in the agent deserves a positive evaluation.[107] The dispositions of one's cognitive faculties seem at least in part constitutive of one's epistemic character, and as such, reliabilist judgments are arguably a kind of epistemic character judgment.

Whichever way we look at it, non-accidentality or expression of character, moral responsibility seems analogous to epistemic justification *by the externalists' own lights*. Of course, as externalists, we don't just look at the cognitive faculties in isolation. We consider the fit between cognitive functioning and the environment.

6.5 Externalism Loud and Proud

Taking a step back, I think careful consideration of this point helps us identify why the externalist's attempt to craft a distinct notion of epistemic blame that captures internalist intuitions is so fraught. At the heart of internalism is the following idea:

---

reasons but isn't consciously malicious isn't blameworthy. Rather, it is just that they are less blameworthy that the person who is maliciously unresponsive to the right moral reasons (the person who knows the right moral reasons but flouts them). I want to flag that the externalist reliabilist can say something similar. Conscious awareness of the grounds of one's belief can sometimes be an indication that one's belief was reliably formed (insofar, broadly speaking, basing one's beliefs on evidence is a reliable belief forming process). As such, conscious awareness of the (good) grounds of one's belief can mean that, at least in some cases, one's belief is more justified.

[106] Arpaly prefers the discussing "depth of concern" rather than character (2002, 239-242). This is mostly because she thinks that talk of character is looser and less precise than is necessary for her purposes. Though I agree with the distinctions she draws, I don't think those details are necessary for my purposes.

[107] Questions about cognitive character and reliabilism will be discussed further below. See section 7.

Autonomous Reason

> The target of epistemic evaluation is autonomous reason. When we evaluate epistemic subjects and their beliefs, we must isolate what is distinctly them.[108] It is unfair to evaluate a subject with reference to factors that are outside of her control. To evaluate someone is to put someone on the hook for something, and it is wrong to put someone on the hook for things they have no control over.

Externalism involves rejecting this idea. In trying to craft a theory of epistemic blame that accommodates internalist intuitions, I worry that externalists think their denial of the above claim is a timid concession. It is not. Is a central, robustly defended commitment of externalism. To the committed externalist, no interesting or significant kind of epistemic normativity is a function of only how well a subject is doing by her own lights.[109] A central externalist tenet is that any purely internalist epistemic norm is disconnected from what is of ultimate epistemic value: the truth. In arguing that epistemic normativity requires considering the subject's external environment, it is true that externalism acknowledges that there is an element of luck. Factors that are outside the epistemic subject's control will play into our epistemic successes and failures, and as such our epistemic evaluations. However, this isn't to open the door to an unwanted guest; it is to recognize that this is the way things are. It is to acknowledge that there is no plausible way to isolate the individual from her environment and end up with something that is epistemically significant.

This should inform how the externalist thinks about brain in vat cases, and Cohen's new evil demon. Internalists tend to look at these cases and say we should consider how we would treat an internal duplicate in the real world. Given we have the intuition that the real-

---

[108] Nagel gave the following characterization of this powerful intuition: "when we blame someone for his actions we are not merely saying it is bad that they happened, or bad that he exists: we are judging *him*, saying he is bad, which is different from him being a bad thing" (1976, 322).

[109] For example, Greco has argued that externalism doesn't just amount to rejecting internalist account of justification, but any purely internalist epistemic evaluation: "Internalism is false as a thesis about any interesting or important sort of epistemic evaluation, and any corresponding sort of epistemic normativity" (2005, 5).

world internal duplicate is justified, we should think the brain in a vat is as well; to do otherwise would be unfair. But to make that move is to say that what is internal is epistemically significant, and consequently that epistemic evaluation and normativity is about whether one is doing well by one's own lights.[110] The committed externalist should see this as an odd move, and an incorrect characterization of what is going in the brain in a vat case. The pull to give the brain in the vat epistemic credit isn't pumped by the idea that they think they are doing well, or that they are satisfying epistemic standards as they understand them. What is pumping the intuition is that they would have true beliefs if they behaved in an epistemically identical fashion in the typical environment (the real world).

Consider Thomas Nagel's discussion of moral luck (1976). His focus is the moral analogue of Autonomous Reason: the idea of the goodwill isolated from external forces. The claim that either moral or epistemic evaluation must target the person, what is within their control, rather than their environment has serious intuitive pull. However, as Nagel argues, our moral judgements do not seem to merely track the isolated person. Rather, they take into account external forces and circumstances.

> What has been done, and what is morally judged, is partly determined by external factors. However jewel-like the goodwill may be in its own right, there is a morally significant difference between rescuing someone from a burning building and dropping him from a twelfth-story window while trying to rescue him…[What we do is] limited by the opportunities and choices with which we are faced, and these are largely determined by factors beyond our control. (323)

[110] I appreciate that some internalists may think that all of this is too hasty, and that internalism doesn't necessarily amount to the thesis that epistemic normativity is about coherence or grounded in whether one is doing epistemically well by one's own lights. However, it should be noted that it is a central aspect of many prominent version of internalism (access internalism in particular). See, e.g., Descartes 2017, Bonjour 1985, Chisolm 1989, Foley 1987. In particular I would imagine that evidentialists may try and avoid this characterization of their brand of internalism. However, it is not clear to me that the characterization can be avoided. Perhaps they would try to appeal to a notion of proper basing: a belief is justified so long as an epistemic subject properly bases her belief on her evidence. But any attempt to characterize proper basing as something other than "properly based by the agent's own lights" will make it the case that the view is no longer internalist.

He terms this phenomenon "moral luck": instances "where a significant aspect of what someone does depends on factors beyond his control, yet we continue to treat him in that respect as an object of moral judgement" (323). The epistemic analogue is clear. Forces and circumstances outside of an epistemic subject's control impact the particular beliefs she holds and whether those beliefs are reliably formed. Moreover, these external factors impact what her epistemic character is like. Importantly, the externalist doesn't take the presence of epistemic luck to undermine epistemic normativity. Rather, the externalist understands it as part of the nature of epistemic normativity. Epistemic normativity isn't fundamentally isolating and evaluating believers as separate from their environment: it is about evaluating the fit between the functioning of our cognitive faculties and our environment.

Particularly for naturalistic externalists, this commitment ought not be a timid concession. When discussing the pervasiveness of moral luck, the way in which it is impossible to escape, Nagel writes:

> The inclusion of consequences in the conception of what we have done is an acknowledgement that *we are parts of the world*, but the paradoxical character of moral luck which emerges from this acknowledgement shows that we are unable to operate with such a view, for it leaves us with *no one to be*. (328; emphasis mine)

To be a naturalistic externalist is to claim that epistemic normativity and our lives as epistemic subjects are parts of the natural world, not merely constructed etiquette.

Where has this discussion led us regarding the plausibility of the externalist pluralistic framework? I have argued that externalists can, at best, endorse a norm of epistemic responsibility that is contextually lucid. While this will allow for the accommodation of some internalist verdicts (e.g., the dogmatic climate change denier), it will not allow for the incorporation of all internalist verdicts (e.g., Norman the clairvoyant). Moreover, I have argued that in analyzing how the externalist pluralistic framework would handle various cases, we are

confronted with an awkward result: the analogies used to motivate the pluralistic framework don't seem to draw the parallels between ethics and epistemology in the right places. On further inspection, it seems that externalist accounts of epistemic justification are already analogous to accounts of moral responsibility that share similar psychological commitments. Finally, I argued that in appreciating that this is the right place to draw the parallel, we are confronted with the idea that, given the externalists' fundamental commitments, it is not clear that they should want anything like an internalist-friendly notion of epistemic blame that attempts to isolate the subject from external forces. It is in this insight that I think we find the root of the issues associated with trying to defend this kind of pluralistic epistemic framework. Maybe, the best thing to do would be to abandon it all together.

7. Anticipating an Objection: Process Pluralism

I do want to consider one last attempt an externalist might make to save a pluralistic framework. I could imagine an externalist suggesting the following: We grant that positing an internalist norm of epistemic responsibility is inconsistent with our fundamental commitments. We grant that we will never be fully able to accommodate all internalist intuitions. That said, surely, we can make some room for the intuitions prompted by cases like the dogmatic climate change denier and the person trapped in the pernicious echo chamber – cases in which our intuitions are pulled in different directions. We can do that by positing a notion of epistemic blame that tracks an externalist norm concerning epistemic character. This would stand in contrast to strict and straightforward process reliabilism, the norm that governs epistemic justification.

What would an externalist norm concerning epistemic character look like? Something like John Greco's "agent reliabilism" might do the trick (1999). To be clear, Greco does not construct agent reliabilism as a theory of epistemic blame. Rather, he proposes agent reliablism

as a theory of the kind of justification that is necessary for knowledge; it is "a general framework for any adequate epistemology" (293). In short, agent reliabilism constrains the kinds of processes that are relevant to determining one's epistemic status. The relevant processes are those that are tied to subjects' cognitive characters. According to agent reliabilism, "knowledge and justified belief are grounded in stable and reliable cognitive character. Such character may include both a person's natural cognitive faculties as well as her acquired habits of thought" (287). One of Greco's central defenses of agent reliabilism is that it can account for the internalist intuition that knowledge requires subjective justification, not merely reliability. Agent reliabilism can be understood, in some important sense, as trying to track the individual – what can be properly attributed to her, rather than to mere luck or chance.[111]

> The dispositions that a person manifests when she is thinking conscientiously are stable properties of her character, and are therefore in an important sense hers. Accordingly, in an important sense a belief that satisfies [the conditions posited by agent reliabilism] … will be subjectively appropriate—it will be well formed from the point of view of the person's own character and motivations. Even more importantly, this kind of subjective appropriateness captures the sense in which knowers must be sensitive to the reliability of their own evidence. Namely, evidence that generate knowledge does so in a way that is grounded in the knower's cognitive character; specifically, the character she manifests when she is motivated to believe the truth. (290)

Insofar as knowers must be sensitive to the reliability of their evidence, the view is externalist and escapes the clutches of skepticism. Insofar as the epistemically significant processes are tied to cognitive character, the view accords with the intuition that we ought only evaluate the person.[112] Finally, insofar acting in character requires acting by own's own motivations, the

[111] This is a clear example of an externalist theory that is trying to have its cake and eat it too.

[112] Of course, it is worth noting that, in true externalist fashion, this view does completely isolate the individual from her environment, chance, and luck. This follows straightforwardly from the fact that the reliability component of the view is externalist. Greco also writes that, on agent reliabilism, cognitive "character may include both a person's natural cognitive faculties as well as her acquired habits of thought" (1999, 287). Insofar the natural cognitive abilities we have are the result of chance, the view entails that epistemic evaluation is open to the influence of external forces (not just that which is in the subject's control).

view captures the idea that justification requires that we are acting in light of, rather than merely conforming to, norms (i.e., subjective justification).

At the outset, I think it is worth noting that this view is really evidence for my earlier claim that for both internalist and externalists, epistemic justification is best understood as analogous with moral responsibility, praise, and blame. That aside, we could consider recasting the view not as a theory of epistemic justification, but rather as a theory of epistemic blame, while maintaining straightforward and simple process reliabilism as our theory of epistemic justification.

> *Agent Reliabilism as a Theory of Epistemic Blame*
> A person S is epistemically blameless with respect to their belief that *p* so long as *p* results from stable and reliable dispositions that make up S's cognitive character. S is epistemically blameworthy with respect to *p* so long as *p* results from stable but unreliable dispositions that make up S's cognitive character.

What are the advantages of doing this? It would allow us to give evaluations like the following. Not only is the dogmatic climate change denier unjustified insofar as her belief-forming process is unreliable, but she is also epistemically blameworthy given that unreliability is grounded in a stable vicious disposition in her character. This evaluation is significant, because it is possible to be unjustified but epistemically blameless, as in the case of the person trapped in a pernicious echo chamber. This person is also unjustified insofar as her belief-forming processes are unreliable, but she is epistemically blameless given the reliability is grounded primarily in her environment, rather than as a stable vicious disposition of her cognitive character. We can see this by considering that she would form true beliefs if she was a different environment—not in an echo chamber, or in an epistemically "good" echo chamber).

The externalist's argument for going this route would be this: we recognize that epistemic normativity arises from considering both (1) our cognitive function and (2) our external environments. That said, we can in some sense locate the "primary" source of

(un)reliability in one of those two places. When it comes to unjustified belief, if we can locate the primary source of unreliability in the subject's cognitive character, the subject is blameworthy with respect to the belief. If we can locate the primary source of unreliability in the environment, then she is blameless.

We should be clear that this doesn't amount accommodating all internalist intuitions. What was said above about the case of Norman the clairvoyant holds here too; there is no such thing as blameworthy justification on this view. In short, this is just an account that posits there are different ways of failing the epistemic justification condition: blameworthy unjustified, and blameless unjustified. Additionally, we should be clear that this framework of epistemic blame isn't tracking the internalist notion of epistemic responsibility. Given the externalist commitments of the view, the claim isn't that cognitive character is the product of agential control and therefore the more appropriate target of epistemic evaluation. Remember, "[t]he dispositions that a person manifests when she is thinking conscientiously are stable properties of her character, and are therefore in an important sense hers", though these dispositions are in part the result of her innate abilities, environment, etc. (Greco 1999, 290). But perhaps this is good enough. After all, it explains the force of our intuitions in the dogmatic climate change denier case, and the intuition that there is something meaningfully different about that kind of unreliability, and the kind of unreliability that is present in an echo chamber.

I think there are reasons to be wary of this approach. It is not clear to me, if we are drawn to agent reliabilism, why we would posit it as a theory of epistemic blame as opposed to theory of epistemic justification (as was intended by Greco). In other words, if we are drawn to agent reliabilism, are we really creating a more illuminating framework by positing it as distinct from, and in addition to, process reliabilism? This goes back to my earlier point:

externalist theories of justification are best understood as theories of epistemic responsibility. A lack of epistemic justification is already analogous to moral blameworthiness.

In addition, once we are committed externalists, it is not that there is an epistemically significant distinction between unreliability that is primarily located in the environment as opposed to the epistemic character of the subject. When it is located in either place, we are granting that it can be out of the subject's control, and up to chance. Recall that in agent reliabilism, the move to focus on the epistemic cognitive character of the subject doesn't seem motivated by the idea that this is normatively significant because one is in control of one's cognitive character.[113] As such, why think that there is an interesting or significant normative difference when unreliability is located in cognitive character? Consider that it is not clear in cases where the unreliability is primarily located in cognitive character; changing the external environment might be the correct remedy (i.e., change the social stakes that are causing dogmatic climate change denier's motivated reasoning). As argued above, part of externalism is that it understands subjects and their epistemic lives as objects and events. Once we grant this, it is not clear that there can be a principled distinction between unreliability located in character as opposed to the environment.

Maybe there are satisfying answers to both of these concerns. Ultimately, I understand this as in-house debate between reliabilists: a debate about whether there is only one or multiple epistemically significant and interesting ways to carve up processes. Of course, the pluralistic externalist framework I have been describing in this section will only work if we hold the latter view. It should be noted however that if we go in for this latter view, many will

[113] Therefore, agent reliabilism not capturing this idea that some argue is central to moral blame. Some argue that "unlike judgments of causal responsibility, blame says something about the person qua moral agent" (Coates and Tognazzini 2012, 200). It is not clear that agent reliabilism would be able to capture another feature some argue is necessary to blame: its "particular moral force" (Hieronymi 2004).

likely want to argue that simple process reliabilism and agent reliabilism aren't the only epistemically significant types of processes. Once we allow this, I think we will struggle to find principled answers to the question of which processes concern blameworthiness, which justification, and which determine yet unnamed evaluations.

Consider Neil Levy and Mark Alfano's claim that "some of our most significant epistemic achievements" are instances of intergenerational social epistemic success they term "cumulative culture", and epistemic phenomena in which information "is not merely transmitted but which then becomes a platform for the next round of social innovation" (2019, 6). Levy and Alfano's key examples of epistemic success resulting from cumulative culture are cases in which humans are able to successfully navigate and flourish in challenging and diverse environments, hereafter, cases of "ecological success" (7). They ask us to consider situations in which staple food crops are toxic when not prepared properly: the Nardoo plant in South Australia or cassava in South America (8). Indigenous communities in these regions developed complex preparation techniques to rid these food sources of toxins and make them safe for consumption. Levy and Alfano think that attributing the epistemic success involved in detoxifying these plants to individual members of these indigenous communities is likely inaccurate because such ecological problems are resistant to being identified and solved by individual, independent reasoners.[114] Levy and Alfano argue that this kind of ecological epistemic success results from an innate human disposition that is, at the individual level, an

[114] For example, the ecological problem of plant toxicity is resistant to being solved by independent reasoners because it is challenging to identify the connection between the food source and the symptoms of toxicity:

> The connection between the plants that must be detoxified and the ill health, for example, is opaque; only some people develop the symptoms of cyanide poisoning [when they consume cassava], for instance, and then only after a prolonged period of reliance on cassava. Individual cognition is ill equipped to detect such distal, stochastic causal regularities. (Levy and Alfano 2019, 9)

How, then, are these ecological problems solved, i.e. how do humans come to have ecologically successful behaviors?

epistemic vice: extreme overimitation (11).[115] Humans demonstrate a strong "conformist bias" (10). In other words, "a disposition to copy the behavior and acquire the beliefs of the majority relatively unselectively and unreflectively" (10). Levy and Alfano's claim seems to be that these ecological problems get solved gradually and overtime, without any of the individual community members' awareness that they are participating in a diachronic, socially-distributed problem-solving activity. The knowledge that results from this activity is preserved despite the lack of individual, conscious understanding because of the disposition to unreflectively imitate. What they take this to show is that some of the most impressive human epistemic feats are laudable only at the collective level. They are epistemic successes for which no individual(s) deserve credit.

Instead of cashing out cumulative culture in terms of epistemic virtues and vices, we could instead take cases like these as evidence that there are multiple interesting and epistemically significant ways to carve up processes. The thought would be that it can be epistemically significant to look at both long- and short-term processes. But if the pluralistic framework makes room for these considerations, it is not clear that this leaves the notion of epistemic blameworthiness in particularly good standing. First, it will likely be difficult to give a principled argument for saying that some processes track epistemic blameworthiness while others do not. Second and more importantly, once we open the door to this degree of process pluralism, why should we think that epistemic blameworthiness is an appropriate notion? Accepting this degree of process pluralism is a vindication of the idea that environment and agent-external forces are deeply relevant to epistemic normativity. Remember, epistemic blameworthiness is arguably trying to appease the intuition that there is something normatively

[115] Levy and Alfano claim that although they are using virtue epistemology scaffolding, they "do not assume the truth of virtue epistemology in any of its forms" (2019, 2). Namely, they are not committed to the virtue epistemologist's analysis of knowledge, nor to any particular accounts of the nature of epistemic virtues.

significant about isolating (to the best of our ability) the agent from her environment and outside forces. These ideas are in tension with one another.

8. Epistemic Blaming as Instrumentally Valuable

It may seem like the work of this paper has been largely deflationary. I have been arguing that the recent work on epistemic blameworthiness has failed to account for how this notion intersects and overlaps with the standing literature on epistemic justification. My view is that theories of epistemic blame will collapse into theories of epistemic justification, whether we are internalist or externalist. In addition, I've contended that conceptual analysis of epistemic blaming will not be philosophically illuminating. In stepping back and considering this chapter's contribution, I can imagine that some may argue that I am failing to account for the obvious: epistemic blame is a significant and robust part of our epistemic lives. We engage in practices of epistemic blaming and debate whether interlocutors are epistemically blameworthy on the day-to-day (though maybe not in so many words). How can I acknowledge this while also minimizing or eliminating the role of epistemic blame in our epistemological theorizing?

First, whether one interprets my arguments as aiming to eliminate the notion of epistemic blameworthiness from our epistemic theorizing depends on how much one thinks is necessarily built into any notion of blameworthiness, and how much is up for revision. Let me explain. One way of understanding my argument is as saying that we already have theories of epistemic blameworthiness, they just haven't been appropriately labeled. On this reading, I am not eliminating epistemic blameworthiness from our epistemic theorizing. I think this route is particularly plausible if one has already committed to access internalism before engaging with my arguments. I say this because traditional accounts of access internalism already use the language associated with folk accounts of moral blame (e.g., responsibility,

accountability, etc.). That said, some might argue the following: externalist theories of justification are too dissimilar from our folk accounts of blame in general to be considered theories of epistemic blame. Take philosophers like Pamela Hieronymi (2004) and Susan Wolf (2011) who argue that blame has a kind of force or affective bent. Externalist theories of epistemic justification can't accommodate an epistemic analogue, given the room these theories make for the influence of luck. In light of this, some might want to argue that, at least when it comes to my arguments about externalism, I am suggesting that we be eliminativists about epistemic blameworthiness. Given the scope of this project, I leave it to the reader to adjudicate between these options. After all, one's previously held epistemological commitments will come to bear, and my aim in this chapter isn't to settle the debate between internalism and externalism. Rather, it is to give an account of how epistemic blame interacts with these fundamental epistemological commitments.

That said, I do want to argue for the following: if one is an externalist (as I have shown myself to be in the preceding chapters), one ought not be discouraged by the idea that externalism is incompatible with a theory of epistemic blameworthiness. First, as I will show below, this position is compatible with positing that epistemic blaming is philosophically interesting and important. Moreover, I think there is room for arguing that the eliminativist attitude towards epistemic blameworthiness is an insight, rather than a concession. As was discussed previously, a fundamental commitment of externalism is the idea that epistemic normativity is about the fit between our epistemic faculties and our environment. It seems appropriate that full appreciation of this idea might involve setting epistemic blameworthiness to the side. After all, an externalist should argue that epistemic success, and positive epistemic status, can't be achieved by the individual alone. We can't, as individuals, always consciously reason ourselves into justified beliefs. In fact, sometimes when we consciously reason, we

might lose justification. Abandoning epistemic blameworthiness could allow for greater appreciation of the practical insight of externalism: epistemic success does not merely require cultivating traits and dispositions in individuals, it requires creating healthy epistemic environments and communities.

The above arguments do not entail that we should have a philosophically deflationary attitude towards epistemic blaming. Though I think that a conceptual analysis of our practices of epistemic blaming will not be philosophically fruitful, that is not to say that I don't think epistemic blaming isn't important or philosophically interesting. On the contrary, practices of epistemic blaming are integral to epistemic success. Consider the discussion of dialogical deliberation in Chapters 1 and 2. Hugo Mercier and Dan Sperber's interactionist theory of reasoning posits that our epistemic capacity for conscious reflection evolved for, and is most successful in, deliberative dialogical contexts (2017). Epistemic blaming is an essential part of productive deliberative dialogue. In epistemically blaming our interlocutors, in holding them epistemically accountable, we are requiring them to share reasons with us and thereby allowing evidence to be presented for public consumption and scrutiny.[116] Moreover, if we accept Mercier and Sperber's argument, epistemic blaming is not only prompting our interlocutors to share their reasons with us, but also triggering the conscious production of reasons in our interlocutor. Practices of epistemic blaming, of questioning one another's epistemic responsibility, are integral to the back-and-forth of epistemically productive dialogue in which conscious, reflective reasoning thrives. Epistemic blaming is instrumentally essential to our

[116] Some might be uncomfortable with this claim, on the grounds that blaming can prompt a kind of defensiveness that can cause an interlocutor to "close off", shut down, or refuse to engage in conversation. But here I think it is important to remember that our practices of blaming are very diverse and need not be bombastic. Blame can be gentle. Indeed, part of being a good epistemic community member might involve knowing when blame should be leveled gently or harshly.

epistemic success, even if the practices do not track a genuine notion of epistemic blameworthiness.

This section is all to say, my arguments are consistent with acknowledging that epistemic blaming practices and discussions of epistemic blameworthiness are regular and robust parts of our lived epistemic experiences.

9. Conclusion

This chapter has explored the notion of epistemic blame. It has argued that there is no real room for a theory of epistemic blame that is distinct from justification on either the internalist or externalist frameworks. If one is an internalist, then the person trapped in the pernicious echo chamber outlined at the chapter's outset has justified beliefs. This is just to say that she is epistemically blameless. If one is an externalist, then this person has unjustified beliefs. The committed externalist doesn't need to cushion this verdict by claiming that she is epistemically blameless. This is because committed externalists believe that epistemic normativity is about the fit or relationship between our cognition and environment. Our epistemic lives are always subject to influence from external forces, and cases like this highlight that commitment. Finally, I have argued that though conceptual analysis of epistemic blaming will be philosophically unilluminating, more instrumental accounts of the diverse practice can be philosophically rich.

In closing, I want to comment on the association between social epistemology and epistemic blame. There are a couple of reasons for this connection. First, epistemologists interested in epistemic blaming take themselves to be investigating what is (at least paradigmatically) an interpersonal phenomenon. Second, some epistemologists try and cash out epistemic blameworthiness by appealing to epistemic obligations we have to each other.

In either case, the thought is that epistemic blame is part of our interpersonal epistemic lives and therefore within the view of social epistemology.

Stepping back and thinking about this chapter in relation to the whole dissertation, I want to suggest that epistemic blameworthiness and conceptual analyses of epistemic blaming ought not be the focus of social epistemology. I argued in Chapter 2 that internalism can't account for the interactive interpersonal epistemic dependence that is a necessary and important feature of our epistemic lives. Social epistemology needs externalism. But there is no real room for a robust notion of epistemic blame on the exernalists framework. Social epistemologists might be concerned with this result. After all, given the interpersonal dimensions of epistemic blame outlined in the paragraph above, it seems like the kind of phenomenon that they should care about. I think this concerned is misplaced. There is still room for philosophically illuminating discussion about the instrumental value of epistemic blaming. More importantly, social epistemologists should remember that robust intuitions about epistemic blame assume that there is something like an epistemic agent that can be isolated from her external environment. This is arguably at odds with social epistemology's central project: to account for the ways in which social forces influence our epistemic lives. The cases that tempt us to reach for explanations regarding epistemic blameworthiness and blamelessness should serve as reminders that epistemic normativity is a function of cognition-environment fit. Deeper appreciation of this could be the prompt we need to shift (or at least expand) the focus of epistemology as a whole. We ought not only care about individual cognitive dispositions that enable reliable reasoning in isolation. Epistemic success requires healthy epistemic environments and communities.

BIBLIOGRAPHY

Arpaly, N. and Schroeder, T. 1999. 'Praise, Blame and the Whole Self.' *Philosophical Studies*, 93: 161-188.

Arpaly, N. 2002. 'Moral Worth.' *The Journal of Philosophy*, 99(5): 223-245.

———. 2003. *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford, England: Oxford University Press.

Audi, R. 2006. 'Testimony, Credulity, and Veracity.' In J. Lackey and E. Sosa (eds), *The Epistemology of Testimony*. New York, NY: Oxford University Press.

Ballarini, C. 2022. 'Epistemic Blame and the New Evil Demon Problem.' *Philosophical Studies*.

Byrne, R. M. 1989. 'Suppressing valid inferences with conditionals.' *Cognition*, 31(1): 61-83.

Bonjour, L. 1980. 'Externalist Theories of Empirical Knowledge.' *Midwest Studies in Philosophy*, 5: 53-73.

———. 1985. *The Structure of Empirical Knowledge*. Cambridge MA: Harvard University Press.

Boult, C. 2017a. 'Epistemic normativity and the justification-excuse distinction.' *Synthese*, 194: 4065-4081.

———. 2017b. 'Knowledge and Attributability.' *Pacific Philosophical Quarterly*, 98: 329-350.

———. 2019. 'Excuses, exemptions, and derivative norms.' *Ratio*, 32(2): 150-158.

———. 2021a. 'The significance of epistemic blame.' *Erkenntnis*

———. 2021b. 'Epistemic blame.' *Philosophy Compass*, 16(8): e12762.

———. 2021c. 'There is a distinctively epistemic kind of blame.' *Philosophy and Phenomenological Research*, 103(3): 518-534.

Brandom, R. 1998. 'Insights and Blindspots of Reliabilism.' *The Monist*, 81(3): 371-392.

Brown, J. 2020a. 'What is epistemic blame?' *Noûs*, 54(2): 389-407.

———. 2020b. 'Epistemically blameworthy belief.' *Philosophical Studies*, 177(12): 3595–3614.

Carruthers, P. 2011. *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford, England: Oxford University Press.

Chisholm, R. 1989. *Theory of Knowledge*. Englewood Cliffs, NJ: Prentice Hall.

Coates, D. J., Tognazzini, N. A. 2012. 'The Nature and Ethics of Blame.' *Philosophy Compass*, 7(3): 197-207.

Cohen, S. 1984. 'Justification and Truth.' *Philosophical Studies*, 46(3): 279–295.

Collin, F. 2019. 'The Twin Roots and Branches of Social Epistemology.' In M. Fricker, P. J. Graham, D. Henderson, N. J. L. L. Pedersen, *The Routledge Handbook of Social Epistemology*. Routeledge Handbooks Online: Routledge.

Comesaña, J. 2010. 'Evidentialist Reliabilism'. *Noûs*, 44(4): 571-600.

Conee, E. and Feldman, R. 2004. *Evidentialism: Essays in Epistemology.* Oxford, England: Oxford University Press.

Craig, E. 1990. *Knowledge and the state of nature: an essay in conceptual synthesis*. Oxford, England: Clarendon Press; New York: Oxford University Press.

Dennett, D. C. 2015. *Elbow Room, New Edition: The Varieties of Free Will Worth Wanting*. Cambridge, MA; London, England: MIT Press.

Descartes, R. 2017. *Meditations on First Philosophy: with selections from the Objections and Replies* (second edition). Cambridge, England, Cambridge University Press.

Evans, J. St. B. T. 1989. *Bias in Human Reasoning: Causes and Consequences*. Hillsdale, NJ: L. Erlbaum Associates.

Fantl, J. 2015. 'Human Knowledge/Human Knowers: Comments on Michael Williams' 'What's So Special about Human Knowledge?'.' *Episteme*, 12(2): 269-273.

———. 2019. 'Evidentialism as an Historical Theory.' *Australasian Journal of Philosophy*, 98(4): 778-791.

Foley, R. 1987. *The Theory of Epistemic Rationality*. Cambridge, MA: Harvard University Press.

Frankfurt, H. 1971. 'Freedom of the Will and the Concept of a Person.' *Journal of Philosophy*, 68: 5-20.

Fricker, M. 2016. 'What's the Point of Blame? A Paradigm Based Explanation.' *Noûs*, 50(1): 165-183.

Gigerenzer, G. 1998. 'Ecological intelligence: An adaptation for frequencies.' In D. Dellarosa Cummins and C. Allen (eds), *The Evolution of Mind*. New York, NY: Oxford University Press.

———. 2011. 'How to Make Cognitive Illusions Disappear: Beyond "Heuristics and Biases".' *European Review of Social Psychology*, 2(1): 83-115.

Gilbert, M. 2004. 'Collective Epistemology.' *Episteme* 1(2): 95-107.

Goldberg, S. C. 2010. *Relying on Others: An Essay in Epistemology*. Oxford, England: Oxford University Press.

———. 2020. 'Epistemically engineered environments.' *Synthese*, 197(7): 2983-2802.

Goldman, A. 1967. 'A Causal Theory of Knowing.' *Journal of Philosophy*, 64(12): 357-372.

———. 1976. Discrimination and Perceptual Knowledge. *The Journal of Philosophy* 73(20), 771-791.

———. 1979. 'What is Justified Belief?' In G. S. Pappas (ed), *Justification and Knowledge: New Studies in Epistemology*. Boston, MA: D. Reidel Publishing Company.

———. 1993. 'Epistemic Folkways and Scientific Epistemology. *Philosophical Issues*, 3: 271-285.

———. 1999. *Knowledge in a Social World*. Oxford: Oxford University Press.

———. 2009. 'Internalism, Externalism, and the Architecture of Justification.' *The Journal of Philosophy*, 106(6): 309-338.

Goldman, A., O'Connor, C. 2021. 'Social Epistemology'. *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/win2021/entries/epistemology-social/.

Greco, J. 1999. 'Agent Reliabilism.' *Philosophical Perspective*, 13: 273-296.

———. 2003. 'Knowledge as Credit for True Belief.' In M. DePaul and L. Zagzebski (eds), *Intellectual Virtue: Perspectives from Ethics and Epistemology*. Oxford, England: Clarendon Press; New York, NY: Oxford University Press.

———. 2005. 'Justification is not internal.' In S. Matthias and E. Sosa (eds), *Contemporary Debates in Epistemology*. Malden, MA: Blackwell Publishing.

Harman, G. 1973. *Thought*. Princeton, NJ: Princeton University Press.

Halberstadt, J. and Wilson, T. 2008. 'Reflections on Conscious Reflection: Mechanisms of Impairment by Reasons Analysis.' In J. E. Adler and L. J. Rips (eds), *Reasoning: Studies of Human Inference and its Foundations*. Cambridge, England; New York, NY: Cambridge University Press.

Hauser, M., Cushman, F., Young, L., Kang-Xing Jin, R. and Mickhail, J. 2007. 'A Dissociation between Moral Judgements and Justifications.' *Mind and Language*, 22(1): 1-21.

Hawthorne, J. and Srinivasan, A. 2013. 'Disagreement Without Transparency: Some Bleak Thoughts.' In D. P. Christensen and J. Lackey (eds), *The Epistemology of Disagreement: New Essays*. Oxford, England: Oxford University Press.

Hieronymi, P. 2004. 'The force and fairness of blame.' *Philosophical Perspectives*, 18: 115–148.

Heller, M. 1995. 'The Simple Solution to the Problem of Generality.' *Noûs*, 29(4): 501-515.

Johnson-Laird, P. N. and Byrne, R. M. J. 2002. 'Conditionals: A theory of meaning, pragmatics, and inference.' *Psychological Review*, 109(4): 646-678.

———. 1991. *Deduction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kant, I. 2014. *Groundwork of the Metaphysics of Morals*. G. Gregor and J. Timmermann (trans). Cambridge, England: Cambridge University Press.

Kelly, T. 2016. 'Historical versus Current Time Slice Theories in Epistemology.' In B. P. McLaughlin and H. Kornblith (eds), *Goldman and His Critics*. Chichester, West Sussex; Malden, MA: Wiley Blackwell.

Korcz, K. A. 2000. 'The Causal-Doxastic Theory of the Basing Relation.' *Canadian Journal of Philosophy*, 30(4): 525-550.

Kornblith, H. 1980. 'Beyond Foundationalism and Coherence Theory.' *Journal of Philosophy,* 77(10): 597-612.

———. 1985. 'Ever Since Descartes.' *The Monist*, 68(2): 264-276.

———. 2002. *Knowledge and its Place in Nature*. Oxford, England: Oxford University Press.

———. 2011. 'Why Should We Care About the Concept of Knowledge?' *Episteme*, 8(1): 38-52.

———. 2012. *On Reflection*. Oxford, England: Oxford University Press.

Kripke, S. A. 1980. *Naming and necessity.* Cambridge, MA: Harvard University Press.

Kuhn, D. 1991. *The Skill of Argument*s. Cambridge, England; New York, NY: Cambridge University Press.

Kuhn, D., Shaw, V. and Felton, M. 1997. 'Effects of dyadic interaction on argumentative reasoning.' *Cognition and Instruction*, 15(3): 287-315.

Levy, N. 2014. *Consciousness and moral responsibility*. Oxford, England; New York, NY: Oxford University Press.

Levy, N. and Alfano, M. 2019. 'Knowledge from Vice: Deeply Social Epistemology.' *Mind* 129(515): 887-915.

Lackey, J. 2006. 'Introduction.' In J. Lackey and E. Sosa (eds), *The Epistemology of Testimony*. New York, NY: Oxford University Press.

———. 2007. 'Why We Don't Deserve Credit for Everything We Know.' *Synthese*, 158(3): 345-361.

———. 2009. 'Knowledge and Credit.' *Philosophical Studies,* 142(1): 27-42.

———. (ed) 2015. *Essays in Collective Epistemology*. New York, NY: Oxford University Press.

Littlejohn, C. (Forthcoming). 'A plea for epistemic excuses.' In J. Dutant (ed), *The New Evil Demon: New Essays on Knowledge, Justification and Rationality*. Oxford, England: Oxford University Press.

Longino, H. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, NJ: Princeton University Press.

———. 2002. *The Fate of Knowledge*. Princeton, NJ: Princeton University Press.

———. 2022. 'What's Social about Social Epistemology?' *The Journal of Philosophy*, 119(4): 169-195.

Lucas, E. J. and Ball, L. J. 2005. 'Think-Aloud Protocols and the Selection-Task: Evidence for Relevance Effects and Rationalizations.' *Thinking and Reasoning* 11(1): 35-66.

McKenna, M. S. 1998. 'The Limits of Evil and the Role of Moral Address: A Defense of Strawsonian Compatibilism.' *The Journal of Ethics*, 2(2): 123–142.

Meehan, D. 2019. 'Is epistemic blame distinct from moral blame?' *Logos and Episteme,* 10(2):183-194.

Mercier, H. and Sperber, D. 2011. 'Why do humans reason? Arguments for an argumentative theory.' *Brain and Behavioral Sciences* 34(2): 57-111.

———. 2012. 'Reasoning as a Social Competence.' In H. Landemore and J. Elster (eds), *Collective Wisdom: Principles and Mechanisms*. Cambridge, England; New York, NY: Cambridge University Press.

———. 2017. *The Enigma of Reason*. Cambridge, MA: Harvard University Press.

Mill, J.S. 2000. *Utilitarianism*. ProQuest Ebook Central: Electronic Book Company.

Millikin, R. G. 1987. *Language, Thought, and other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.

Moody-Adams, M. M. 'Culture, Responsibility, and Affected Ignorance.' *Ethics*, 104(2): 291-309.

Nagel, T. 1976. 'Moral Luck.' Reprinted in R. Shafer-Landau, 2013, *Ethical Theory: An Anthology*. Malden, MA: Wiley-Blackwell.

Nguyen, T. C. 2020. 'Echo Chambers and Epistemic Bubbles.' *Episteme*, 17(2): 141-161.

Nisbett, R. and Ross, L. 1980. *Human Inference: Strategies and shortcomings of social judgement*. Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, R. and Wilson, T. 2002. 'Telling More than We Can Know: Verbal Reports on Mental Processes.' *Psychological Review*, 84(3): 231-259.

Pereboom, D. 2001. *Living Without Free Will*. Cambridge, England: Cambridge University Press.

———. 2014. *Free Will, Agency, and Meaning in Life*. New York, NY: Oxford University Press.

Perkins, D. N. 1985. 'Postprimary education has little impact on informal reasoning.' *Journal of Educational Psychology* 77(5): 562-571.

Piattelli-Palmarini, M. 1994. *Inevitable Illusions: How Mistakes of Reason Rule Our Mind*. New York, NY: Wiley.

Pinker, S. 1997. *How the Mind Works*. New York, NY: W. W. Norton and Company.

Piovarchy, A. 2021. 'What do We Want from a Theory of Epistemic Blame?' *Australasian Journal of Philosophy*, 99(4):791-805.

Resnick, L. B., Salmon, M., Zeitz, C. M., Wathen, S. H., and Holowchk, M. 1993. 'Reasoning in Conversation.' *Cognition and Instruction,* 11(3-4): 347-364.

Rettler, L. 2018. 'In Defense of Doxastic Blame.' *Synthese: An International Journal for Epistemology, Methodology and Philosophy of Science,* 195 (5): 2205-2226.

Rips, L. J. 1994. *The Psychology of Proof: Deductive Reasoning in Human Thinking*. Cambridge, MA: MIT Press.

Robison, J. 2020. 'Moral Worth and Consciousness: In Defense of a Value-Secured Reliability Theory.' *Ergo*, 7(9): 277-305.

Scanlon, T. M. 2008. *Moral dimensions: Permissibility, Meaning, Blame.* Cambridge, MA: Belknap Press.

Sher, G. 2006. *In Praise of Blame.* Oxford, England: Oxford University Press.

Shieber, J. 2019. 'Socially Distributed Cognition and the Epistemology of Testimony.' In M. Fricker, P. Graham, D. Henderson, and Nikolaj Jang Pedersen (eds), *The Routeledge Handbook of Social Epistemology*. New York, NY: Routledge.

Silva, P. 2015. 'On Doxastic Justification and Properly Basing One's Beliefs.' *Erkenntnis*, 80(5): 945-955.

Smart, P. R. 2018. 'Mandevillian intelligence.' *Synthese*, 195(9): 4169-4200.

Sosa, E. 1991. *Knowledge in perspective: selected essays in epistemology*. Cambridge, England; New York, NY: Cambridge University Press.

———. 2007. *A Virtue Epistemology: Apt Belief and Reflective Knowledge*. Oxford, England: Clarendon Press; New York, NY: Oxford University Press.

———. 2015. *Judgment and Agency*. Oxford, England; New York, NY: Oxford University Press.

Srinivasan, A. 2020. 'Radical Externalism.' *Philosophical Review,* 129(3): 395-431.

Stanovich, K. E. and West, R. F. (2008). 'On the failure of cognitive ability to predict myside and one-sided thinking biases.' *Thinking and Reasoning*, 14(2): 129-167.

Stanovich, K. E., West, R. F., and Toplak, M. E. 2013. 'Myside bias, rational thinking, and intelligence.' *Current Directions in Psychological Science,* 22(4): 259-264.

Strawson, P. F. 1974. 'Freedom and resentment.' In P. F. Strawson (ed), *Freedom and resentment and other essays*. London, England: Methuen.

Tomasello, M. 2014. *A natural history of human thinking*. Cambridge, MA: Harvard University Press, 2014.

Tversky, A. and Kahneman, D. 1973. 'Availability: A heuristic for judging frequency and probability.' *Cognitive Psychology,* 5(2): 207-232.

———. 1983. 'Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement.' *Psychological Review*, 90(4): 293-232.

Twain, M. 2009. *The Adventures of Huckleberry Finn*. New York, NY: Penguin Books.

Wason, P. and Evans, J. St. B. T. 1975. 'Dual Processing in Reasoning.' *Cognition* 3(2): 141-154.

Williams, M. 1997. *Groundless Belief: An Essay on the Possibility of Epistemology*. New Haven, CT: Yale University Press.

———. 2004. 'Is Knowledge a Natural Phenomenon?' In R. Schant (ed), *The Externalist Challenge*. Berlin, Germany; New York, NY: Walter de Gruyter.

———. 2008. 'Responsibility and Reliability.' *Philosophical Papers*, *37*(1): 1-26.

———. 2015. 'What's So Special about Human Knowledge?' *Episteme*, 12(2): 249-268.

Williamson, T. 2000. *Knowledge and Its Limits*. Oxford, England: Oxford University Press.

———. (Forthcoming). 'Justifications, excuses, and sceptical scenarios.;' In J. Dutant (ed), *The new evil demon*. Oxford, England: Oxford University Press.

Wolf, S. 2011. 'Blame, Italian Style.' In R. J. Wallace, R. Kumar, and S. Freeman (eds), *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*. Oxford Scholarship Online: Oxford University Press.

Zagzebski, L. T. 1996. *Virtues of the mind: an inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge, England; New York, NY: Cambridge University Press.

———. 2012. *Epistemic authority: a theory of trust, authority, and autonomy in belief*. New York, NY: Oxford University Press.