UNIVERSITY OF CALIFORNIA

Santa Barbara

On the Nature and Ethics of Belief

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Philosophy

by

David Anthony King

Committee in charge:

Professor Daniel Z. Korman, Co-Chair

Professor Aaron Zimmerman, Co-Chair

Professor Elinor Mason

December 2024

The dissertation of David Anthony King is approved.

_____

Elinor Mason

_____

Daniel Z. Korman, Committee Chair

_____

Aaron Zimmerman, Committee Chair

December 2024

On the Nature and Ethics of Belief

ACKNOWLEDGEMENTS

If I have accomplished anything of value here, it is only thanks to the many people who have taken the time to talk and think we with me over the years. And it is everyone who has cared enough to help my thinking along over the years that I most want to thank.

I especially owe a huge debt to Aaron Zimmerman and Dan Korman. Dan Korman has taught me more about philosophy than any other person in this world. Aaron Zimmerman has been a huge help in pushing my thinking forward, sharpening my skills, and developing expertise in my main field of research. Thanks so much to both of you.

What follows, in no particular order, is a list of some of the many people who have left a lasting impact on my thinking and helped me to develop into the philosopher I am today. Thanks to Elinor Mason, Matthew Lund, Ásta, Isabelle Peschard, Nathan Salmon, Bas van Fraassen, Paolo Mancosu, Thomas Barrett, Tom Holden, Matt Hanser, Ellen Miller, David Landy, Alice Sowaal, Patrick Skeels, Rick Lamb, Stephen Troxel, Yunzhong Mao, Isabelle Brady, Chaya Haugland, James (Wangjin) Xie, Jon Charry, Enri Lala, Natalie Ries, Griffin Baxley, Sneha Jiju, Tom Costigan, Sean Pierce, Sam Zahn, Daniel Story and Chris Britton.

David King

602 N 3rd St. #7, Lompoc, CA, 93436 • (805) 280-5307 • dking20@calpoly.edu

RESEARCH

Areas of Specialization

- Philosophy of Mind (esp. Belief, Emotion, and Neurodiversity)
- Cognitive Science (esp. Belief, Emotion, Meaning, CSR, and Neurodiversity)
- Applied Ethics (esp. Ethics of Belief and Emotion, Neuroethics)

Areas of Competence

- Ethics and Social Philosophy
- Meaning in/of life
- Metaphysics
- Epistemology
- Philosophy of Language

EDUCATION

University of California, Santa Barbara — Santa Barbara, CA
Ph.D., Philosophy — 2024
Dissertation: *The Nature and Ethics of Belief*
Chairs: Aaron Zimmerman and Dan Korman

San Francisco State University — San Francisco, CA
M.A., Philosophy — 2017
Thesis: *Epistemic Modality*
Chair: Ásta (Sveinsdóttir)

Rowan University — Glassboro, NJ
B.A., Philosophy — 2014
Honors: Summa Cum Laude, Senior of Distinction Award

PUBLICATIONS AND FORTHCOMING WORK

Forthcoming

"Three Theories of Belief" with Aaron Zimmerman, in *Belief,* eds. Eric Schwitzgebel and Johnathan Jong, Oxford University Press

"Making Sense of Neurological Differences," in *AJOB Neuroscience*

Under Review

"Beliefs without Believers"
"Crises of Meaning," with Yunzhong Mao

In Preparation

"Taking Infinite Beliefs Seriously" (Dissertation Chapter)
"The Belief-Emotion Nexus" (Dissertation Chapter)
"Inconsistencies of Belief and Behavior" (Dissertation Chapter)
"Neurodiversity and Neuroprejudice in the Study of Mind"

PRESENTATIONS

"The Philosophy of ADHD" – *UCSB ADHD Graduate Group* (Invited, but date TBD)
"Neurodiversity and Neuroprejudice in the Study of Mind" – *UCSB Philosophy Graduate Colloquium,* May 2024
"The Shifting Sands of Belief" – *Cal Poly Philosophy Research Workshop,* April 2024
"The Shifting Sands of Belief" – *UCSB Philosophy Graduate Colloquium,* April 2024
"Believing the Victim" – *UCSB Philosophy Graduate Colloquium,* May 2022

TEACHIING EXPERIENCE

California Polytechnic State University (2024-Present) San Luis Obispo, CA

Lecturer (Duties: Lecture, Course Design and Office Hours)

- Knowledge and Reality (x1), Ethics and Political Philosophy (x2)

University of California, Santa Barbara (2017-2024) Santa Barbara, CA

Lecturer (Duties: Lecture, Course Design and Office Hours)

- Philosophy of Mind (x2), Intro to Philosophy (x2)

Teaching Assistant (Duties: Leading Discussion Sections, Grading, and Office Hours)

- Lower Div Course: Intro to Philosophy (x6), Intro to Ethics (x3), Critical Thinking (x2)
- Upper Div Courses: Philosophy of Mind (x1), Ethics (x1), Theory of Knowledge (x2), Metaphysics (x1), Philosophy of Science (x1),

Ethics Bowl Coach (2023-2024)

San Francisco State University (2014-2017) San Francisco, CA

Lecturer (Duties: Lecture, Course Design, and Office Hours)

- Critical Thinking (x4), Intro to Philosophy (x1), Intro to Philosophy and Religion (x2), Great Thinkers: East and West (x1)

Teaching Assistant (Duties: Grading and Office Hours)

- Symbolic Logic (x1), Modern Philosophy (x1), Philosophy of Risk (x1)

HONORS & AWARDS

Paul Wienpahl Award for Excellence in Teaching, UCSB Philosophy Department	2024
Ralph W. Church Community Service Award, UCSB Philosophy Department	2024
Graduate Division Dissertation Fellowship	2023
Senior of Distinction Award, Rowan University	2014

LEADERSHIP & OUTREACH

Coached the UCSB Ethics Bowl Team (placed 2nd in Nationals)	2023-2024
Mentored 10 student research projects through UCSB's RMP Program	2021-2024

REFERENCES

Aaron Zimmerman
Affiliation: UC Santa Barbara, Philosophy Department
Titles: Professor, Department Chair
Relationship: Advisor, Dissertation Chair
Email: aaronzimmerman@ucsb.edu

Daniel Korman
Affiliation: UC Santa Barbara, Philosophy Department
Title: Professor
Relationship: Advisor, Dissertation Chair
Email: dkorman@ucsb.edu

Elinor Mason
Affiliation: UC Santa Barbara, Philosophy Department
Title: Professor
Relationship: Dissertation Committee Member
Email: elinormason@ucsb.edu

Abstract

On the Nature and Ethics of Belief

by

David King

The aim of this dissertation is to make substantive progress toward understanding the nature and ethics of belief. The job of understanding belief is too large a task for a single dissertation, but each chapter bites off a small portion of the larger task, each doing a job that has largely been neglected by philosophers of belief but, which is important for understanding the nature and ethics of belief. The dissertation starts by exploring the question of whether we should say we have an infinite or finite number of beliefs. Many philosophers have gone on record on the issue, but few have developed arguments or unpacked the implications of answering the question one way or another. I make the case that any plausible account of mental architecture should be able to render the verdict that we have an infinite number of beliefs. I then proceed to explore the relationship between the belief relation talked about by philosophers of language and beliefs. I make the case that, contrary to popular opinion on the matter, beliefs are not profitably understood as relations. I then turn to the question of what level of consistency we should expect between beliefs and behavior. A number of thinkers in philosophy, psychology, and cognitive science more broadly have used inconsistencies between agents' stated beliefs and their behaviors to argue for the existence of novel mental phenomena or to make the case that many of the

most paradigmatic cases of belief are not belief at all—that the religious do not generally believe in the tenets of their religion, that the deluded to do not believe their delusions, and that many racial egalitarians actually unwittingly hold racist beliefs. I make the case that they mostly have no case, as belief and behavior are rarely so in sync that we should be ever be surprised by someone acting in a manner inconsistent with their beliefs. Living according to one's beliefs is an achievement; it is not constitutive of belief. I then turn to the question of how our beliefs and emotions interact. I argue that what we believe is often dependent upon what we feel to an extent that has gone underappreciated in the literatures on emotion and belief. I call the locus of this interplay, *the belief-emotion nexus,* and utilize the model of it developed herein to make progress on a variety of thorny issues in the philosophy and cognitive science of belief and emotion. Finally, I turn to the issue of believing the victim. Many claim that we have some special duty to believe the testimony of those reporting to be victims of sexual assault. I explore the issue with a careful eye toward the nature and ethics of belief issues involved and make the case that, under at least some common understandings of what "believe the victim" means, it is in fact rational to do so.

TABLE OF CONTENTS

Introduction

The papers in this volume are concerned with the nature and ethics of belief, broadly construed. All research has to start by taking some assumptions for granted, and I take the existence of beliefs and the fact that beliefs are objects of moral significance to be about as obvious as anything in philosophy can be. So, the work contained herein proceeds by taking these two things for granted.

Among the motivating thoughts for this project was the idea that it is important to study these two topics together. I have been unmoved by the arguments of skeptics on these issues. And I regard with suspicion any account of the nature of belief that cannot be reconciled with a satisfying view of doxastic morality. I regard with comparable suspicion any account of doxastic morality that has no place for a sophisticated and nuanced account of the nature of belief. Understanding doxastic morality as I see it requires understanding the nature of belief and understanding the nature of belief requires understanding how it is that beliefs can be objects of moral evaluation.

The papers in this volume do not amount to a successful holistic account of the nature and ethics of belief. They do however make substantive progress toward that goal—each chapter serving as a substantial lemma on the way to future and more developed views on the nature and ethics of belief. They offer constraints on what such a view can and should look like, while casting serious doubt on many popular and influential ideas on these matters that have been defended or taken for granted in the literature. They are early dispatches from a rapidly developing cartography, which has yet to successfully map its territory, but which has nonetheless produced substantial results that will need to be taken into account by any future cartographies setting out to map the same territory. What is more, they each address

areas of significance for the nature and ethics of belief that have mostly been neglected in relevant philosophical literature. They are important groundwork for any serious study of the nature and ethics of belief to take into account, even if they leave many important questions unsettled.

The contents of this volume lean much more heavily on issues relating to the nature of beliefs. But even when discussing the metaphysics, cognitive science, or epistemology of belief, the ethical issues are often not very far off, even if they are not being engaged with directly. Chapter 3, for example, lays the groundwork for thinking about the ethics of belief attribution and for undermining many popular views about attributions of racism and the possession conditions for racist beliefs. So, even when the papers are not directly about ethics issues, for many of them there is still an important sense in which they really are all about the ethics of belief. Many of the topics here were chosen for their relevance to ongoing debates within the ethics of belief.

Each chapter is written as a self-contained document. So, the chapters can be read in any order. I have ordered the papers so that the volume starts with issues related to the nature of belief and slowly transitions to matters that have greater import for the ethics of belief as it proceeds. So, there is some added utility in reading them in the order they are included in the volume, as the contents of each chapter are more apt to be of relevance to chapters that are nearby. But each chapter can be understood on its own.

I have opted not to bog this introduction down with unnecessary citations, as most of the relevant citations can be found in the chapters described. Citations only appear here when I move outside of the content of the chapters to either discuss literature that is not mentioned

in a body chapter or make a content attribution to a work that is not made in the body chapters, even if that work is in fact cited.

This introduction will provide an overview of the contents of the dissertation and offer a narrative that makes clear how the chapters relate to each other and to wider discussions on the nature and ethics of belief. It will thus not only summarize the work, but explain its significance to broader issues, which may not be addressed within the chapters themselves.

1 – The Metaphysics of Belief: Chapters 1 and 2

Chapter 1 of the volume is focused on the issue of whether we should say that we have an infinite number of beliefs or a finite number. Having to explain how it is that we could have an infinite number of beliefs would amount to a serious constraint on any theory of belief and many theories of belief are incompatible with the possibility. So, establishing to what extent we should think having an infinite number of beliefs is a success condition for a theory of belief is important for settling what theories of belief to take seriously.

A number of theorists have weighed in on the issue in print, but few have developed substantive arguments in one direction or the other. For most theorists, the matter is settled by other theoretical commitments they hold for other reasons, like respect for natural language practices, or to their broader picture of the metaphysics of the mind or mental architecture. There has been little investigation on the matter that proceeds by arguing from premises that are likely to be widely acceptable, nor much investigation into the implications of going one way or the other on the question.

At issue are a number of other questions about how we should theorize about the mind and where language fits in. Those who wish to best respect our natural language attribution

practices think beliefs should be sentence or proposition-individuable, as they appear to be when we use natural language to attribute them to one another. Following respect for these practices to its natural end seems like it should lead us to the conclusion that we have an infinite number of beliefs. This is usually defended by appeal to examples, like the following: you believe the number of suns in the solar system is one. This means you believe it is not two, three, four, etc—ad infinitum. For every number x, such that x is not equal to 1, it seems intuitively plausible to say that you believe the number of suns in the solar system is not x and that there is a distinct belief for every admissible value of x.

Skeptics about infinite beliefs tend to be motivated by what they regard as a satisfying picture of mental states or of mental architecture. Identity theorists (e.g. Smart 1959 and Armstrong 1968), for example, want every mental state to be identifiable with a brain state, and no principled account of the nature of brain states seems likely to allow an infinite number of discrete brain states to be realized by a finite amount of matter. Those who are not identity theorists do not require mental states to be identifiable with brain states, but still are likely to require all states of the mind to be explicable by a finite amount of brain matter, even if the story of how the mental relates to the physical is not a simple story of identifying mental states with specific brain states. And insofar as they want to say that our folk psychological kinds, like beliefs, desires, etc. are mental *states*, it seems as though they have to have some account of what mental states are such that there is a distinct mental state for every distinct belief. How there could be an infinite number of discrete mental states realized by a finite number of brain states is not altogether clear. Thus, many attempts to take seriously the metaphysics of folk psychological kinds like beliefs runs into the challenge of saying what the states could be such that there could be an infinite number of

them. Many, motivated by such considerations, are ready to buck the evidence from gleaned from examination of our natural language use and attribution practices and say that we have a finite number of beliefs, that this is required for integrating the theory of belief with cognitive science, and that saving the appearances of natural language practices is not as relevant of a theoretical consideration as the things they care about—namely, a satisfying account of mental architecture and integration of a theory of belief with the science of the mind (e.g. Quilty-Dunn and Mandelbaum 2018).

I enter the debate in a somewhat strange way, motivated by similar motivations as those that are bringing skeptics to their position, but arguing for a different conclusion. No satisfying picture of mental architecture, I argue, should be unable to produce the result that we typically have an infinite number of beliefs. This is because any picture of mental architecture that requires finite beliefs is unlikely to be efficient or productive enough to make sense of our memory systems or the productivity of language; nor is it likely to have evolved. It is a trivial matter for a computational system to house an infinite amount of information (in the sense that an infinite number of distinct sentences with distinct contents can be retrieved from the system), so long as information storage and retrieval is computational in its own right. The picture I sketch as a more realistic picture of mental architecture is one wherein storage processes are aimed at making information optimized for search and retrieval. This means minimizing the amount of information that has to be combed over, optimizing its organization for searchability, eliminating redundancies, and having memory systems that can draw basic inferences, to leverage a finite store of information into an infinite quantity of information being retrievable. This means that information is likely to be stored in ways that does not respect the boundaries of sentences

or propositions, and sentences that spring to mind as remembered information are not copied or moved from memory stores, but constructed on the fly, tailor made for whatever behavioral or computational purpose they are needed for. The result is that we have an infinite number of beliefs that can be brought to mind at moment's notice, despite a finite amount of brain matter. This information springs to mind as remembered information that is known. The information that the number of suns is not, four, five, eighty-seven, etc. is all information that springs to mind just as easily, without any conscious act of inference being performed.

This raises questions about what beliefs actually are, on my picture—something I am not yet prepared to fully weigh in on and which I do not delve into in this dissertation. On my picture, there are no entities, like representations, modes of representation, or mentalese sentences that persist through the life of a belief. This is not to say that such entities do not exist—it is just to say that none of them persist through the life of a belief and thus that none of them can be numerically identical with nor a necessary condition for the existence of a belief. Beliefs have a distinct modal profile from all such entities. On my picture, it seems better to say that an infinite number of beliefs are realized by a finite number of states, but that the beliefs themselves are not identifiable with mental states. There is not a mental state for every belief.

Given that the claim that beliefs are mental *states* is taken to be definitional by most philosophers, this may seem like a problem. An easy "fix" for this "problem" might be to draw on the tradition within the philosophy of language studying belief. This tradition studies belief by studying belief reports—natural language sentences of the form *S believes that p.* This tradition often says "belief is a relation." While it is not clear that everyone in

this tradition would agree to the statement that beliefs *just are* relations, many would—and it is those that would who have a view that is worth considering here. If we view beliefs as relations to abstract entities, like propositions, then it seems that an easy fix for the inadequate count of brain states would be to take the infinite number of relations to an infinite number of propositions to be what the relevant states, and thus the beliefs, are. We thus have an infinite number of them on this propositional attitude account and no problem of saying what a mental *state* could be, such that a finite amount of brain states could give rise to an infinite number of mental states, with distinguishable states for each belief.

While this may be a tempting route to go, I do not think that the account is defensible. One of the foremost reasons for my hesitation is that it seems as though we have a non-uniqueness problem on this account. Any Platonistic account of mental states as relations to abstract entities has the problems of saying which entities we are related to, how we are related to them, and how it is that we either are only related to those ones or despite being related to many kinds it is only the relations to those particular ones that are our mental states. I do not believe these desiderata can be satisfied. The full argumentation for this cannot be explored here. But, in brief, any plausible account of how it is that we manage to be connected to one of these abstract entities seems sufficient to connect us up to a wide variety of abstracta. And if that is the case, then there are too many relations to too many relata for any choice to say "these here are beliefs" to be non-arbitrary. And should the arbitrariness be embraced, there is then no reason to say that beliefs are in any way important to the study of the mind, because there is nothing particularly special about the

relation such that it, rather than some other relation to some other abstracta, could be constitutive of mentality.[1]

This non-uniqueness problem is not explored here. I explore it in work not included here. While I cannot rehearse all of the arguments I have against extant attempted solutions to the problem, as doing so would require more space than I can afford the issue here, I do explore a second reason for doubting the relational account in chapter 2. One of the relata in the belief relation is supposed to be an agent. And I make the case in chapter 2 that beliefs have persistence conditions that are very different than the belief relations we uncover by examining belief reports, because they can persist through changes in agents, while belief relations cannot. If this is right, it is sufficient to show that the belief relations we learn about in studying belief reports cannot be beliefs. Beliefs have a different modal profile. Belief relations may be useful in uncovering the existence of beliefs or communicating information about them, but they cannot be identical with beliefs.

So, if in the life span of a belief, there is no representation, mode of presentation, mentalese sentence token, or any other form of mental furniture that persists through the life of the belief; and if no account of beliefs as relations between agents and abstracta is plausible, what kind of things can beliefs be? I do not have the answer to that question. I have said here a great deal about what kinds of things beliefs cannot be. But finding the optimal account of the metaphysics of beliefs to make sense of infinite beliefs without reducing them to relations will have to be the project of future research.

[1] No one has written on the non-uniqueness problem for propositional attitudes, but the non-uniqueness problem has been explored for propositions themselves (e.g. Moore 1999 and Armour-Garb and Woodbridge 2010) and for numbers (e.g. Benacerraf 1965 and Balaguer 1998). Non-uniqueness problems and questions about their significance are also common in other disciplines (Science Direct has lit reviews available on non-uniqueness problems in a variety of subject areas.)

2 – Belief and Action: Chapter 3

The idea that some property or relation having to do with action is partially or wholly constitutive of belief goes back at least as far as the classical pragmatists, like Bain and James. There are, however, a multitude of ideas about just what belief's relation to action is in the belief literature. Though it's not often acknowledged, theorists disagree about the consequences of the fact that belief has some kind of intimate relation to action. At minimum, most thinkers are prepared to believe that it is a (or the) function of belief to guide action. A wide range of thinkers take things even further, expecting that there should be some kind of consistency between belief and action—that how one acts is a definitive test for what one believes, such that failing to act in accordance with one's beliefs is tantamount to not believing. Such thinkers have used belief-action inconsistencies to argue that religious people do not believe in the tenets of their faith, that the deluded do not believe their delusions, that many racial egalitarians are actually unwitting racists, and that people do not actually believe the strange things they purport to believe. Call such cases *belief denials.*

While I take no issue with the claim that it is a function of belief to guide action, I take issue with the idea that we can expect a high degree of consistency between belief and action. Chapter 3 of the paper weighs into the debates wherein a wide range of thinkers in philosophy and cognitive science have utilized belief-action inconsistencies to argue for belief denials (and often for the existence of novel mental kinds that are distinct from but similar to beliefs). I make the case in chapter 3 that they mostly have no case. Belief-action inconsistencies are common and unavoidable. More disturbingly, they often exhibit a kind of systematicity, wherein agents can *consistently* fail to bring information to bear on

circumstances where its content is relevant in a way that they are blind to. Bringing information to bear on tasks where its content is relevant is a complex and difficult task, one that is especially difficult to achieve in time sensitive decision making and one that can be affected by surprising things, like the adequacy of our mental architecture for handling certain contents and our history of cognitively managing similar situations. The result is that we can expect belief-action inconsistencies to be par for the course. Acting consistently with one's belief is an achievement; it is not constitutive of what it is to have beliefs. It may be a function of beliefs to guide actions, but successfully executing such a task is easier said than done. What beliefs exist to do is not always predictive of what can be done with a belief. Failing to recognize this is tantamount to admitting that we have few or no beliefs. We simply cannot earn a belief denial by appeal to a belief-action inconsistency.

3 – Belief and Emotion: Chapter 4

Chapter 4 explores the interplay between belief and emotion—something I call *the belief-emotion nexus*. I argue that what we believe is often dependent on what we feel—to an extent that has often gone underappreciated in the literatures on belief and emotion. The idea that emotions bias our thinking in certain ways is a commonplace—especially in psychology and affective neuroscience. But exploration of how different emotions affect cognition and the function of doing so is something that has seen insufficient exploration in any of the various literatures on emotion or belief. There is especially very little philosophical work on the matter.

I advance here an error-management theory of emotions, which aims to predict the biasing effects of emotions by appeal to the management of costly errors and provides a

general framework for thinking about the ways in which belief formation, retention, and revision strategies are shaped by emotion. So understood, it is the chief function of emotions to steer us away from costly errors—especially in time sensitive decision making and belief formation—so as to ensure that the worst possible outcomes are avoided. This does not mean that problems are always avoided thereby. Sometimes, attempting to avoid one error steers us right into another. And because evolution is a satisficer and not a maximizer, there is no reason to think that our emotions are infallible in matters of error avoidance. But emotions do put their thumb on the scales to bias the way we interpret the world, affecting how we form beliefs and make decisions, by making us more susceptible or resistant to certain beliefs and choices. They are in effect, like the bumper rails at the bowling lane: they may not result in bowling a good score, but they are consistently good at helping a large enough percent of the species avoid the worst possible scores.

The EMT account of emotions predicts how each emotion will shape our thinking, resulting in a general account of how belief formation, revision, and retention strategies are constantly shifting as a result of shifting emotions. The result of so much biased thinking seems apt to pose a threat to us when too many biased beliefs lead to a confused view about the world. I offer a theoretical framework for discussing the problem and posit a two-tier error avoidance strategy as common to neurotypical humans. Tier one is aimed at the avoidance of costly errors, while tier two is aimed at revising beliefs for consistency and accuracy, when agents are downregulated.

I also make a contribution to the literature on why people believe weird things by sketching out the ways the belief-emotion nexus can contribute to things like radicalization and extremist beliefs. I argue for the existence of a kind of self-reinforcing cycle, which I

call EB-Cycling, which results from failure to downregulate and engage in error correction causing runaway error pileups. In practice it might look something like this: S is angry, so she is more susceptible to believe p. The proposition p, however, if true, warrants more anger, which shapes further belief formation and decision making, and further entrenches p. As new beliefs form, they are systematized, become entrenched, and warrant more anger. As more decisions are made in anger, the tendency to search for more defenses for one's angry beliefs becomes habit. The cycle of anger, anger-biased belief-formation, belief-justified anger, and so on can lead people to some radical beliefs, resulting in agents who may be prone to believe some rather implausible things and to believe that violence is a warranted response on much weaker evidence than they otherwise would, had they not EB-cycled so thoroughly. If we want to understand why people belief such crazy things, we need to start attending to how they are feeling. And insofar as we want to remedy their bad beliefs, we are apt to need to attend to their feelings more than the cogency of our arguments against them.

5 – Believing the Victim: Chapter 5

Chapter 5 dives into a thorny practical ethics question for the ethics of belief—namely, of whether or not we have some special moral obligation to believe the testimony of those who claim to be victims of sexual assault and whether doing so can be seen as rational. A number of advocacy groups have formed around advancing the position that we ought to believe those claiming to be victims of such crimes. The question then is what is it that such advocacy asks of us and how does it intersect with the rationality of belief? Are we being asked to believe victims in absence of evidence, in the face of countermanding evidence, or

because the evidence favors believing? Is this a problem of investigators and the public at large failing to meet their obligations to be rational agents or are they are obligated to be irrational to fulfill this supposed duty? In order to settle such questions, we need first to examine to what extent it is rational to believe the complainants in such cases—those claiming to be victims.

Of course, the rationality of believing testimony in any individual case is apt to change as the evidential situation for each case changes in the course of the investigation. So, there is not likely to be a verdict about whether it is rational to believe those claiming to be victims of sexual assault generally. But we can ask whether it is rational to *start by believing*, as some advocates claim, in cases where all we have is the contradictory testimony of the complainant and the accused. I make the case that in rape investigations in our contemporary social, economic, and political circumstances, it is rational to start by believing. This does not mean that our starting situation could not be different in such a way as to change the rationality of starting by believing reports of sexual assaults. It also does not mean that our ecology could not change in such a way as to change the rationality of starting by believing. But in our present circumstances, it is far more rational to start by believing the report, all else being equal.

This of course raises further worries that starting by believing the report will unduly bias the investigation toward the complainant and against the accused—something that might be claimed to be immoral or that might be claimed to be at odds with our justice system. But I do not argue that one is morally obligated to believe what is rational, only that it is in fact rational to start by believing, so long as one does not have any defeating evidence. All else being equal, it is more likely that the complainant is telling the truth than not. Further, since

beliefs are likely to happen anyway and because the current system is biased in favor of males over females, there is good reason to think that any biases produced in the investigation will mitigate biases that investigators already hold against the complainant.

That said, just because it is rational for us to start by believing, this does not mean we are capable of doing so. Nor does it preclude other positive, non-evidential reasons that favor suspending belief, or a general attitude toward favoring testimony of innocence. I admit at the outset that I do not have the space to address some of the reasons that might be offered against starting by believing. This is an aim for future work.

Rima Basu has taken the position that believing the victim is a case wherein pragmatic encroachment plays a role, resulting in the standards of evidence required for belief being lowered for believing the complainant, due to the moral stakes. I make the case here that pragmatic encroachment is not required to deliver the result that it is rational to believe the victim—and further that pragmatic encroachment does not obviously favor believing the complainants in such cases to begin with. In fact, it is often just the opposite. What is at stake for the accused, if they are wrongly convicted, seems in most cases to be greater than what is at stake for the complainant if they do not get justice. There thus seems to be little reason to suppose that pragmatic encroachment is needed to explain anything here or that it even gives the result that Basu is after. Thus, whether or not we have good independent reasons for accepting pragmatic encroachment, there is little reason to think it in any way illuminating or of service in understanding the morality or rationality of starting by believing.

6 – Conclusion

The five body chapters of this work are loosely connected. There are argumentative threads that tie many of them together and they also connect up with argumentative threads that I pursue elsewhere, but could not include here. But they do not yet amount to a successful holistic account of the nature and ethics of belief.

While they leave many questions about the nature and ethics belief unresolved, they each speak to neglected issues in the philosophical literature and make the case that understanding the nature and ethics of belief requires settling many of the issues they address. Each makes a contribution to ongoing debates and helps narrow the field of possible admissible views on the nature and ethics of belief. They thus constitute important groundwork for any future accounts of the nature and ethics of belief.

I. Taking Infinite Beliefs Seriously

This chapter is concerned with the question of whether we do or can have an infinite number of beliefs. The question has been pursued in two different debates within philosophy. One of them is over a view about the structure of justification called *infinitism,* which claims that for a belief to be justified, there must be an infinite chain of justifying reasons supporting it. One major source of contention over that view is whether the view can be dismissed on the grounds that we do not have an infinite number of beliefs.[2]

The other area of philosophy where the question is important is in the study of the nature of belief. Having an infinite number of beliefs would amount to a serious constraint for any account of the metaphysics of belief to accommodate and would eliminate from consideration a number of compelling views on the grounds that they reduce beliefs to entities that come only in finite quantities. One historically important account of the nature of belief, for example, says belief is a relation between an agent and a mentalese sentence (Fodor 1978), where the relation is instantiated by having such sentences stored in a belief box[3] or given a metarepresentational tag.[4] It does not seem as though we could possibly have an infinite number of distinct mentalese sentences in our heads. So, having an infinite number of beliefs could potentially rule out this account and many like it entirely.

[2] See Williams 1981 and Fumerton 1995 for such critiques. See Klein 1999 for the infinitist response.

[3] This kind of "boxology" approach to mental architecture is often associated with a broadly Fodorian approach to the study of the mind. The first explicit discussion of a belief box comes from Schiffer 1981. Perhaps the most developed defense of the need for a belief box comes in Sperber 1997. Boxology approaches to mental architecture and metarepresentation have been employed quite a bit in philosophical work on the mind (e.g. Nichols and Stich 2004), but to what extent they are essential and what commitments to such boxology even amount to is not very clear. For an example of where such confusions and controversies are important to the metaphysics of belief, see Schwitzgebel 2002 and Quitly-Dunn and Mandelbaum 2018.

[4] "Tagging" is most associated with the work of Cosmides and Tooby, but has become a popular alternative approach to metarepresentation and mental architecture, which would not require a belief box in order to keep track of which representations are believed.

Whether or not we have an infinite number of beliefs thus has serious import for debates about the nature of belief and the structure of justification. The aim of this paper is to examine whether it is plausible to think we have an infinite number of beliefs and explore what import an affirmative answer would have for the study of the metaphysics of belief. I will ultimately argue that the reasons for accepting we have an infinite number of beliefs far outweigh the reasons for denying it, and further, that the resulting picture of mental architecture that emerges is far more appealing than the view of mental architecture that motivate many to deny that we have an infinite number of beliefs.

In section 1, I will start with an initial positive case for the claim that we do have an infinite number of beliefs. In section 2, I will take a look at the arguments from those who are skeptical that we have an infinite number of beliefs (hereafter: *the infinity skeptics*). I will make the case that there really is no good argument against the claim. Rather, there are theoretical approaches to belief that cannot accommodate such a possibility. One's stand on the issue is typically dictated in advance by one's theoretical assumptions about the metaphysics of belief and mental architecture. In section 3, I will turn to some additional considerations that speak in favor of saying that we have an infinite number of beliefs. In section 4, I will turn to considerations governing the optimization of storage and retrieval, arguing that common assumptions about the persistence conditions of mental states are rendered implausible by the constraints of good information architecture. Further, if these principles of good information architecture are reflected in our minds, we can easily make sense of having an infinite number of beliefs. However, this picture of the mind is one that will wreak havoc on many ideas advocated by philosophers of belief.

Section 1 – Tacit Beliefs and The Argument for Infinite Beliefs

At the heart of the question of whether we have an infinite number of beliefs are entities that are often called *tacit beliefs*.[5] I will call also them *tacit beliefs* throughout the paper, but it is important to flag up front that whether they truly are beliefs is part of what is in question in these debates. In calling them tacit beliefs, I am following convention, but no arguments in this paper takes their status as beliefs as assumed common ground.

The basic idea is that tacit beliefs are entities or states that do not have explicit, discrete sentence or proposition individuable representations in the mind.[6] Rather, commitment to their truth is produced via the presence of some other kind of representational entity.

There are certainly some important questions about what we can be said to tacitly believe. For example, a criterion of tacit belief that said we tacitly believe any proposition that is a logical consequence of the beliefs that do have explicit, discrete representation in the mind seems implausible, as it would mean many of us are committed to the truth of many things that we could never hope to figure out, which are nonetheless entailed by our non-tacit beliefs. What is more, it is not unreasonable to think that we have some false beliefs that entail claims that are inconsistent with entailments of our true beliefs, leading to contradictory entailments. And if classical logic is to be believed, this would have the implication that we tacitly believe everything, since everything is entailed by a contradiction. Finally, it seems as though we can reflect on what we already believe, draw deductively valid inferences, and form new beliefs—which would be impossible, were

[5] See Lycan 1986 for a classic exploration of the phenomenon

[6] It is worth flagging that the term "tacit belief" is sometimes used for unconscious representations. That is not how the term is being used here or in the literature that is relevant to these points.

entailment the criterion of tacit belief. So, entailment certainly cannot be a viable criterion of tacit belief.

As to what criteria do determine our tacit beliefs (assuming we have them), that is a question that I cannot hope to pursue in this section (though I will return to it at the end of the paper). Instead, I will offer some examples of tacit beliefs that it seems particularly tempting to report as cases of genuine belief, to me at least. Two of the cases I offer will show that it is tempting to attribute an infinite number of beliefs to agents. The other two will be cases of tacit belief, where a belief denial seems particularly counterintuitive, thus bolstering the case that tacit beliefs genuinely are beliefs. How these cases work together will be explained after I have given the cases.

Call the first case *the infinite series.* To generate the infinite series and motivate its plausibility, let us start with a few observations. I seem to believe (i) that one is greater than zero, (ii) that two is greater than zero, (iii) that the beliefs I just reported with i and ii are different beliefs, (iv) that I believed this before I came up with the example, and (v) that it is unlikely that I had sentences of the form of i or ii in my head prior to coming up with the example.[7] And when I think about it for a moment, it seems to me that these facts generalize to an infinite number of propositions. It seems then that for any number x, such that x is greater than zero, I believe that x is greater than zero. Voila. Infinite beliefs.

Call the second kind of case *free negatives.* Suppose that Steven believes he is in his bathroom. Steven's belief that he is in his bathroom also seems to amount to believing he is

[7] In case (v) is not obvious: it seems redundant for me to have a representation that the natural numbers proceed as 0,1,2,3,4,5… and so on, and to have separate representations for all of the relations stored in memory as distinct sentences (e.g. that 1 is greater than zero, that two is greater than zero, that two is greater than 1, that 2 is less than three, etc.). It is possible that I have distinct sentences in storage for some of these relations, but it seems rather pointless and inefficient to have representations for all such relations.

not located in any location x, such that x is not his bathroom or a place in which his bathroom is located. And this is so whether or not he has explicit, discrete sentence or proposition individuable representations of those other locations or whether he has ever represented such places. If someone calls him and asks him whether he is at work right now, he does not consciously draw an inference, inferring that he is not at work, from the premises "I am in my bathroom" and "my bathroom is not at work"—the latter of which is almost certainly a tacit belief anyway.

Call the third kind of case *bizarre composites.* Chihiro is using an authenticator app on her phone to login to a website. She reads the numbers 6148672 in the app and says them over and over to herself until she succeeds in copying the numbers into her web browser, where they were needed. As soon as she inputs the numbers, she forgets them. It seems tempting to say that Chihiro believes that the required number is 6148672 for the full duration those digits are in her working memory. But the numbers 6148672 were held in working memory and the seven digits there exhausted the full limits of the storage available in working memory. So, where is the belief that the required number is 6148672? If beliefs are sentence individuable, as early Fodor (1978) and Carnap (1932 and 1947) thought, then it would seem as though she did not believe the required number is 6148672, as the numbers never left her working memory, and were never combined with any sentence of that form. If on the other hand, beliefs are proposition individuable, then we may not need to deny that she believes the relevant proposition. But we must accept that the belief was either realized by multiple different representations or by only the representation in working memory, which did not have the form of a sentence or enough structure to uniquely pick out a determinate proposition.

Call the fourth case *arbitrary composition.* Jenny has a mental map of her apartment. When asked where the bathroom is, she points to its location. It seems to be obvious that Jenny believes that her bathroom is where she pointed. But the representation of her bathroom's location was stored in a representation of her apartment (the mental map), that is neither sentence individuable nor proposition individuable.[8] From any location in her apartment, she could easily point you to the bathroom, even if she could not come up with sentences to codify all of the relations between locations in apartment and the bathroom. And yet, it seems accurate to report as the object of her belief a sentence or proposition.

In all four cases, it seems tempting (to me at least) to attribute the relevant beliefs, while denying the presence of entities that are sentence or proposition individuable acting as realizers for the belief. The first two cases provide a case that we have an infinite number of beliefs, while the second two cases provide instances where belief denial seems particularly counterintuitive, but where there can be no discrete entity that is sentence or proposition individuable acting as a realizer for the belief.

The arguments from here are relatively straightforward:

The Infinite Series Argument
(IS1) For every number x, such that x is greater than 0, I have a belief that x is greater than 0, which is distinct from any other belief with a different value for x.
(IS2) There are an infinite number of values for x
(IS3) So, I have an infinite number of distinct beliefs

It may be tempting to deny the argument, especially if one is wedded to a particular theory of belief that is incompatible with infinite beliefs (more on this in the next section). The only way out of the argument is the first premise. I will look at some more possible

[8] See Camp 2007 for exploration of thinking with maps and its import for how philosophers have historically approached the mind.

challenges in the next section. But for now, I just want to explore the possibility of merely denying that we have tacit beliefs to try and get out of the argument.

This is where the *arbitrary composition* and *bizarre composites* cases come in. These are cases wherein belief denial seems particularly embarrassing. Even if one does not like cases like *the infinite series* or *free negatives*, where an infinite number of beliefs are being attributed, it seems as though there is still a strong pull to accept that tacit beliefs exist and genuinely are beliefs.

We can thus preemptively bolster the Infinite Series Argument by appeal to those cases:

The Argument from Embarrassment
(AE1) Chihiro and Jenny have beliefs that are not realized by discrete sentence or proposition individuable representations.
(AE2) If so, then Chihiro and Jenny have tacit beliefs.
(AE3) If Chihiro and Jenny have tacit beliefs, then tacit beliefs exist and genuinely are beliefs.
(AE4) If tacit beliefs exist and genuinely are beliefs, then we have no reason to deny that we have an infinite number of tacit beliefs.
(AE5) We have no reason to deny that we have an infinite number of tacit beliefs.

Cases like *infinite series* and *free negatives* show that our intuitive belief attribution practices suggest attributing an infinite number of beliefs, which do not have discrete sentence or proposition individuable representations in the mind. *Bizarre Composites* and *Arbitrary Composition* show that whether or not we want to say we have an infinite number of beliefs, we still have good reason to accept that we have tacit beliefs anyway. And if we must accept that we have tacit beliefs, then the fact (or purported fact) that we do not have an infinite number of discrete sentence or proposition individuable representations in our mind is not sufficient reason to deny that we have an infinite number of beliefs, since tacit

belief does not require the presence of discrete sentence or proposition individuable representations with the content of the tacit belief acting as a realizer for the belief.

We can thus formulate two challenges based on the above reasoning:

The Tacit Belief Challenge
(TB1) Tacit beliefs are beliefs
(TB2) Tacit beliefs are not discrete sentence or proposition individuable representations stored in the mind.
(TB3) So, beliefs are not discrete sentence or proposition individuable representations in the mind.

The Infinite Belief Challenge
(IB1) We have an infinite number of beliefs
(IB2) We do not have an infinite number of discrete sentence or proposition individuable representations in the mind.
(IB3) So, beliefs are not discrete sentence or proposition individuable representations in the mind.

Section 2 – The Infinity Skeptics

I have just made an initial case that we do have an infinite number of beliefs and that this entails that beliefs are not discrete, sentence or proposition individuable representations. The case is compelling, though perhaps not so much so that we ought to be willing to give up on the theories of belief it would rule out. If we have strong enough positive reasons for accepting the theories of belief that the above arguments would rule out, then that may be sufficient reason to reject the arguments above and accept the counterintuitive belief denials in question. But hopefully it helps to paint a picture of why we might initially be tempted by the claim that we have an infinite number of beliefs.

I will offer some additional considerations in favor of infinite beliefs in sections 3, 4, and 5, but before I turn to them, it will be useful to consider the infinity skeptics' challenges to the claim that we have an infinite number of beliefs.

2.1 – The Argument from Psychofunctionalism

The first challenge of the infinity skeptics is obviously question begging, but it is a methodologically appropriate kind of question begging. The basic idea is that we should be psychofunctionalists in our approach to the metaphysics of belief and that any appropriate truly psychofunctionalist account of belief is likely to reduce belief to entities that are not the kind of entities that tacit beliefs can be.

The case is perhaps best understood with a bit of history, which will explain the force, inertia, and institutional entrenchment of this kind of approach.

Somewhere just past the midpoint of the twentieth century, cracks in the dominant behavioralist paradigm in psychology began to show. In the wake of the fall of the behaviorist hegemony the cognitive revolution got underway.[9] The cognitive revolution involved commitment to a level of explanation between neurology and behavior—the cognitive.

A common idea among the proponents of the cognitive revolution was the analogy of a computer. Many thought of the cognitive/mind as being like the software, with the brain being like the hardware. We would need not just an account of the hardware, but of the software, and the relation between the software and the hardware to truly understand ourselves.

A core commitment of the cognitive revolution is representationalism about the mind. This is the thesis that the mind is essentially a representational device. The particular brand of representationalism that proved most popular and influential was computationalism. The

[9] See Miller 2003 for a brief account of the cognitive revolution from one of its major figures.

brain was thought to be essentially a representational device and thought so conceived was computations over representations.[10]

One of the major figures in the cognitive revolution was Jerry Fodor. Fodor took the project of giving an account of the cognitive (the software) as at its core being a project that would vindicate our folk psychology. Fodor's early position was that the entities of folk psychology were relations between agents and sentences in a language of thought (mentalese sentences) (Fodor 1975 and Fodor 1978). He later opted to change his account, following Nathan Salmon (1986), to taking the relation to be between agent, mode of representation, and proposition (Fodor 1994).[11] But the core of his analysis of belief remained constant: belief is realized by discrete mentalese sentences at the software level—adding to it that they were to be individuated not just by sentences, but propositions[12] as well.

Many in the cognitive revolution have followed Fodor on many or all of these points. Some are not so keen on the language of thought hypothesis, but still take beliefs to be representations that can be individuated by appeal to propositions or natural language sentences. Others take the language of thought hypothesis to still be the best game in town for understanding the mind, but are much more open to accounts of the mind that are at odds with our folk psychology.[13]

Psychofunctionalists typically adopt representationalist analyses of folk psychological states, like belief, taking it as a necessary condition on believing p that some discrete

[10] Fodor 1968 was influential in the spread of this idea.
[11] See also Fodor 2008 for Fodor's LoT views after 30+ years of debate.
[12] Propositions here are conceived of as abstract entities.
[13] See Quilty-Dunn, Porot and Mandelbaum (forthcoming) for a spirited defense of the claim that it is still the best game in town.

representation that is sentence or proposition individuable that can be identified with the belief that p be realized in the mind. But psychofunctionalists (e.g. Porot and Mandelbaum 2020) typically take the matter of what the best account of folk psychological states like belief are beyond that to be a largely empirical matter, to be brought out by doing the science of the mind.

Typical to this kind of approach is a commitment to the claim that the our folk psychological terms for mental kinds have natural kind semantics.[14] Finding out what beliefs are is thus not a job for mere intuition pumping but is instead a matter of letting the science of the mind reveal the natural kind that most closely accords with our pre-theoretic notion of belief and then letting reference magnetism take its course, sticking the word "belief" to those entities.

Thus, the argument from psychofunctionalism is that we already know that whatever entities cognitive science may uncover (which ought to be more or less sentence individuable and have persistence conditions that more or less correspond to what we take the persistence conditions of beliefs to be), there cannot be an infinite number of them. Thus, we have to grant that we cannot have an infinite number of beliefs.

This is, to my mind, the most compelling reason to deny that we have an infinite number of beliefs. This is a tradition of work on the mind that has been fruitful and is seen as closely integrated with cognitive science, which takes many of these theoretical assumptions as its starting points. And it is not the way of good science to abandon a good theory, just because it has a few counterintuitive consequences.

[14] Lycan 1988 and Stich 1996 are largely responsible for the psychofunctionalist trend toward thinking "belief" has natural kind semantics. See Poslasjko 2022 for exploration of the Lycan-Stich argument.

However, for anyone moved by the considerations levied in the previous section or those in the sections to come, it is ham-fistedly question-begging as a challenge to the arguments. Thus, no one who is moved by them and is taking their compelling nature as a reason to seriously consider whether to accept accounts within the psychofunctionalist wheelhouse will be moved by this challenge.

However, if one has independent reason to think that "belief" has natural kind semantics and that mental architecture is likely to resemble our folk psychology, then this might still amount to sufficient reason to reject the claim that we have an infinite number of beliefs. The prize of a good theory may be worth the cost of some counterintuitive consequences.

I will argue however, in section 4, that it really is not as good a theory as it is often taken to be.

2.2 – Dispositions to Believe

Robert Audi has been among the most vocal infinity skeptics.[15] Audi thinks that what is going on in the intuition pumps of section 1 is that we are getting the facts wrong, mistaking something that is belief-like for an actual belief.

The belief-like entity in question is a *disposition to believe*.[16] The term *disposition to believe* here is being used as a name. When one believes p, one is disposed to believe p as well. But this is not a *disposition to believe* in the sense that Audi is talking about. Audi uses the term to talk about cases where one merely has the disposition, but does not have the

[15] I take Audi as my main interlocutor here. He is not the only infinity skeptic, and in fact most of the points he makes others have made before him. But he does offer the best formulations of the infinity skeptics' arguments. So, I will note as best as I can throughout where Audi has gotten his ideas, but I will keep Audi as my interlocutor and infinity skeptic representative throughout.

[16] Audi first defends this claim in Audi 1994. But long before him, others used the terminology *disposition to judge* to make essentially the same point (e.g. Sellars 1958, de Sousa 1971, Powers 1978, and Richardson 1981).

belief. Thus, as the term is used here, having a disposition to believe p is inconsistent with having a belief p.

Audi's early argument for the position took the form of an inference to the best explanation argument, though he later turned to more positive arguments for the position, in hopes of getting his critics to accept that it really is the best explanation.

Dispositions to believe are not representations, but instead dispositions to believe certain content when that content is represented. The dispositions to believe are individuable in much the way that beliefs are thought to be—by reference to some sentence or proposition—but they are not beliefs themselves, but instead a different kind of entity, not typically represented as distinct from belief in our folk psychological attribution practices.

Dispositions to believe are like beliefs in that, if one were disposed to believe p, and were to be asked about the truth of p, they will readily assent to the truth of p—as if they had believed it all along.

Dispositions to believe are unlike proper dispositional beliefs, according to Audi. The primary way of distinguishing them is by appeal to memory retrieval. Beliefs are entities that are retrieved from memory stores, while dispositions to believe cause newly formed beliefs (Audi 1994).

To say that they are newly formed beliefs may seem a bit question-begging, but it is important to remember that he is offering an inference to the best explanation argument. So, in principle, not much of his case should actually depend on us actually granting that it is a newly formed belief. Giving an inference to the best explanation argument does not require a theory-neutral description of the facts. It requires a compelling description of the facts that utilizes the resources of the theory or explanatory strategy. What everyone should agree on

is that it was a newly formed representation. And on Audi's explanatory framework, it counts as a newly formed belief. Whether we should accept that it actually is one will depend on whether we should accept his explanation.

So far, it is difficult to see why Audi thinks this is a better explanation than the possibility that the information was antecedently believed. While Audi does identify more distinguishing properties, they all boil down to effects stemming from the fact that beliefs are retrieved from memory stores, whereas the disposition to believe results in a new believed representation being formed, and they all rely on the assumption that a new belief is formed when the new representation is formed. Thus, his whole inference to the best explanation argument stems from the differences that result from a new representation being formed, rather than retrieved from memory stores.

In a later paper (Audi 2020), however, he does move beyond the inference to the best explanation argument. The main challenge in that paper comes from the limits of what is representable. The idea is that at some point in the infinite series, we'll reach a number that is so long that I could not hope to ever represent it. It is a number that is too big to think. Since I cannot represent it, I cannot believe it. Let's call the first number in the number line that is so big that representing it in standard form would not be possible to do in a single human lifetime Ψ (Audi 2020). We can formalize Audi's challenge thusly:

The Argument from Representational Limits
(RL1) You cannot represent Ψ
(RL2) If you cannot represent Ψ, then you cannot hold any beliefs about Ψ or any number greater than Ψ
(RL3) So, you cannot hold any beliefs about Ψ or any number greater than Ψ
(RL4) If you cannot hold beliefs about Ψ or any number greater than Ψ, the number of your beliefs is finite.
(RL5) So, the number of your beliefs is finite.

I do not think we should be moved by this argument at all, since the first premise is obviously false. We have represented Ψ seven times already, by calling it Ψ and by designating it with a description. The fact that we cannot represent it in standard form does not mean that we cannot get at it using a description and represent it in a more tractable form. Audi's own descriptions are a way of representing the number. There seems to be no reason to allow that there is a number that is too big to think from the mere fact that it is too big to think in one mode of representation.

Audi offers an additional reason for claiming that we cannot represent this number. The claim is that it is so big that we would not be able to distinguish it from other nearby numbers, or from numbers that have one digit difference somewhere in the long sequence (Audi 2020, p. 546).

First, there is no difficulty in representing nearby numbers we are distinguishing it from. Consider, Ψ+1 and Ψ-1. We have represented these numbers. Second, the fact that we might not be able to distinguish it from other numbers in one mode of representation does not have any impact on our ability to think the number. I cannot distinguish nearby numbers in Russian, but I can think all of the numbers just fine.

Now that we have seen some additional challenges from the infinity skeptics and why they fail, we can return to the initial question of whether the explanation Audi has to offer is really the best explanation. I see no reason to grant that it is and several reason for thinking it is not.

First, let me say that I agree with Audi that we have dispositions to believe and that some things that have been called tacit beliefs in the past are more appropriately thought of

as dispositions to believe.[17] What is more, in those cases, I do not think that norms of belief attribution in any way recommend attributing belief. Thus, in cases that I take to be correctly analyzed as involving dispositions to believe, he is not at odds with the norms of belief attribution or with anything that I feel tempted by in considering the possibility that we have an infinite number of beliefs.[18] And most of the cases Audi considers fall into this category. So, when it comes to most of the cases that Audi discusses, I am in complete agreement with his analysis of them. But I think that even if he is right about all of those cases, we still have positive reasons for thinking we have an infinite number of beliefs.

There is only one case that he actually offers a substantive analysis in terms of dispositions to believe that actually gets in the neighborhood of cases that I think are compelling instances of tacit beliefs.

The one case he discusses that I do take to be somewhat relevant is the case of believing that 98.124 is greater than 98. The case is a case of an inequality, much like those of the infinite series. Audi considers two possible explanations for what is going on in such a case:

> First, I might believe something to the effect that any integer is increased by adding a decimal to it; hence the moment I think of the proposition that 98.124 is larger than 98, I form the belief that it is. Second, I might have in mind some pattern representing comparative numerical sizes, and immediately see the two numbers as fitting it. To be sure, the apparent immediacy may be only temporal and not epistemic. My newly formed belief may, in the first case be *inferential,* since it is based on my general belief about integers and decimals, or in the second, conceptually *mediated,* given the way it is based on a pattern, yet non-inferential (Audi 1994, p. 422).

Thus, for Audi the two possibilities that might apply to the infinite series are that when I form a new representation of a proposition in the infinite series, either I come to believe it via some general belief that is being applied to draw an inference, or some conceptual apparatus automatically delivers the belief non-inferentially.

[17] Many of his sample cases in Audi 1994 fit this description.
[18] Although, in Audi 2020, he balks at some cases I accept as cases of belief.

I feel less temptation to ascribe belief in Audi's case than I do in the case of the infinite series. I do not have any principle I can appeal to as to why that is.[19] But my methodology from the start has been to start with cases where I feel most tempted to ascribe belief, so I will use the case of the infinite series and apply what he has to say to the case of the infinite series.

While someone might arrive at belief in one of the sentences/propositions in the infinite series via inference, I see no reason to suppose that is what is happening with me, unless the inferences are happening unconsciously. I certainly am not aware of making any inferences, nor would I be tempted to appeal to any general belief about numbers to justify my claim about the inequalities. So, insofar as introspection can be thought of as a guide here, it speaks against the inference proposal. The phenomenology of the experience is just simply one of obviousness—as if failure to recognize that the number is greater than zero would amount to a failure to have represented that particular number. What is more, such information springs to mind in much the same way as any other information. And I feel quite convinced that I knew such facts long before I ever represented them. They spring to mind as if recalled, though I doubt many of the sentences in the infinite series have ever previously been represented in my mind, and even among those that have been, it seems strange to think that my mind would be keeping in storage discrete, sentence or proposition individuable representations of them.

It would seem then that Audi's explanation, the one that is supposed to be better than granting that I believe the propositions of the series, despite never having represented most of them, is that some conceptual mediation has caused me to form the belief spontaneously.

[19] Though some of what I have to say in section 5 may make sense of why his case might be different from the numbers in the infinite series.

But this just seems to amount to saying that I formed the representation using concepts and something about that conceptual machinery caused me to believe it, as if I had believed it all along.

But this explanation is just a redescription of mutually agreed upon facts. In such cases, he wants to say that the formation of new representation constitutes the formation of a belief, whereas the person who accepts tacit beliefs wants to say that the formation of a new representation in such cases amounts to producing a belief that was already held by the agent. Audi wants to say that something about the conceptual mediation caused it to be instantly believed, whereas the advocate of tacit beliefs will want to say that something about the conceptual machinery caused it to be recognized as antecedently believed. But everyone agrees that a new representation was formed, that it was conceptually mediated, that the entity produced was a belief and that a conviction that the newly represented information is believed is likely to be present. What is at issue is whether we should say that the agent believed the relevant sentence/proposition prior to the formation of the representation. It would seem then that this has devolved into a merely verbal debate, wherein all parties agree about the facts, but there is disagreement about when the word "belief" applies. We can find no difference of entities or processes at work in the disagreement.[20]

Thus, my assessment of Audi's criticism is that his direct arguments against the claim that we have an infinite number of beliefs fail and what he is left with is a redescription of the facts that simply refuses to use the word "belief" where the norms of belief attribution prescribe attributing belief.

[20] For exploration of whether this is truly just a verbal disagreement, see Chalmers 2011. See also debates surrounding Parfit's remarks about reductionism and personal identity.

It seems then that Audi's case ends up being much like the psychofunctionalist case in the previous subsection. He is antecedently committed to the claim that beliefs are representations and has no room for believing without the presence of discrete, sentence or proposition individuable representation realizing the belief. In the most compelling cases of tacit belief, wherein the norms of belief attribution seem to prescribe attributing belief, he does not really have a challenge. He is rejecting the norms of belief attribution in favor of a theory of belief that can be made sense of by reducing beliefs to discrete, sentence or proposition individuable representations.

While the possibility of a theory that integrates with science of the mind so nicely and respects so many of our intuitions about the persistence conditions of belief may be reason enough to reject the possibility of infinite beliefs, I see little in his case that actually moves beyond Audi's starting assumptions about the nature of belief. The fact that a new representation is produced in cases where information that is tacitly believed is first represented is common ground between those who do and those who do not go in for tacit beliefs. So all of the facts are the same for Audi and the views he is contending with. It is the descriptions of the facts that seem to be at issue. So, there is no sense in which Audi has produced the best explanation, and there is serious reason to doubt he has even produced a different explanation at all.

His arguments thus are dialectically inert. No one who seriously is tempted to accept that we have an infinite number of beliefs should be tempted by them. Nor should anyone who is not tempted think that they have a leg up in the debate by accepting his explanations of what is going on in such cases. At best, they have a description of the facts that accords with their antecedent theoretical commitments.

2.3 – The Argument for Kind Difference

Perhaps I have been too harsh on Audi. Rather than appealing to what he explicitly says or the motivations he has, we should appeal to the fact that he is pointing to over and over again—that with tacit beliefs a new representation is formed, whereas with non-tacit beliefs representations are not formed. Thus, we have a difference, which suggests a difference of kind. If we are attracted to a theory of reference magnetism, then if we have a difference in kind, the word belief picks out the kind he wants, and the other entities are what get a new name.

We can formalize the reasoning like so:

The Argument for Kind Difference
(KD1) With tacit beliefs, new representations are formed, but with standing dispositional beliefs, no new representation is formed.
(KD2) If KD1, then tacit beliefs and standing dispositional beliefs are entities of a different kind.
(KD3). If tacit beliefs and standing dispositional beliefs are entities of a different kind, then the word "belief" picks out the standing dispositional beliefs and not the tacit beliefs.
(KD4) If the word "belief" picks out the standing dispositional beliefs and not the tacit beliefs, then "tacit beliefs" are not really beliefs.
(KD5) So, "tacit beliefs" are not really beliefs.

I take it that KD4 should be a matter of no-contest for anyone who buys the previous three premises. KD3 is motivated by a theory of "belief" having natural kind semantics. While I am not sure that I buy the claim about "belief" having natural kind semantics, I will not challenge it here.[21]

[21] For exploration of whether belief does have natural kind semantics, see Polsajko 2022. For some classic approaches to belief that did not presume belief to have natural kind semantics (and argued via that route for eliminativism), see Churchland 1981 and Stich 1983.

I see no reason to grant KD2 and much of what I have to say in section 3 will amount to a serious challenge of the claim, but for now I will leave it alone and challenge KD1. It is just plain false that standing dispositional beliefs cannot have new representations formed. First, consider case 4, with Jenny's mental map of her apartment. If Jenny puts the content of some feature of her map into language, she creates a new representation of something she believed before. But she does not believe anything new. If we insist on calling the new mode created a new belief, that can grant Audi that the new representation is a new belief, but it does so in a way that does not undermine the claim that the information was antecedently believed despite having no explicit representation as a discrete, sentence or proposition individuable entity.

It is also the case that we can sometimes figure something out anew faster than we can recall information. For example, I have the times tables for many numbers memorized up to 20. If asked what 13x19 is, I should be able to remember it with the same efficiency and speed that I can normally conjure. That said, the speed and efficiency I can normally conjure is not always as fast as just doing the calculation afresh. Multiply 13x2 to get 26, add a zero and subtract 13 to get 247. I can do that pretty quick. That process may go far more quickly for me than the recall process. Memory retrieval processes can sometimes be slower than just solving a problem anew.

Memory retrieval processes can also sometimes be slower than just looking. Consider: John is asked what color Janey's dress is. If given a moment he would be able to say what it is by recalling the info from memory. But instead, he instinctively looks at Janey's dress in order to answer the question. This does not mean he would not have recalled the info if given a moment to do so. It just means that engaging recall processes is not always the

quickest means to produce the information. We can sometimes get the information via a quick look, despite the fact that we could have recalled it.

Depending on one's theory of belief and the objects of belief, it might also be possible to form new representations of already believed propositions that are already realized by the presence of a discrete sentence or proposition individuable representation. Consider Marcus, who knows that The Rock is Dwayne Johnson. He might believe The Rock is tall, but then produce an occurrent representation of Dwayne Johnson being tall. On some theories of belief (though not all) he is producing the same belief with a different representation.

These are all cases wherein new representations of existing commitments are formed and it seems plausible to think that they are all cases wherein a new representation of already believed information is produced. So, I see no reason to think that the formation of a new representation is particularly telling about whether or not the information was previously believed.

Perhaps then what we need to do is produce a more careful formulation of KD1, in order to cash in on Audi's point. We can try:

> (KD1*) Tacit beliefs are not stored in memory stores, but standing dispositional beliefs are.

Does this new premise save the argument? No. There is no reason for the advocate of tacit beliefs to accept this. Tacit beliefs are stored *in some sense* (more on this in later sections). Tacit beliefs are just not stored as discrete representations that are either sentence or proposition individuable. But the ability to reliably produce representations in cases 1, 2, and 4 seems to be explained by the fact that commitments to the truth of all the relevant propositions is encoded by a mental architecture that is capable of reliably producing the information. So, there seems to be no reason for anyone who accepts tacit beliefs to accept

KD1*. (If this point is not sufficiently clear, it will become clearer in the following sections.)

Perhaps then what we need is a different first premise:

> (KD1**) Tacit beliefs are not stored in memory stores as discrete, sentence or proposition individuable representations, but standing dispositional beliefs are.

Now we have a premise that is most likely to be appealing to both parties in the debate. This will push the debate onto KD2.

In section 3, I will offer some more resources for a denial of KD2, showing how the tacit beliefs have a great deal in common with the non-tacit ones.

In section 4 however, I will throw a significant wrench in the whole debate and offer a view on which *most*, or perhaps even *all* beliefs are tacit. Such a view, if correct, undermines the argument. But it can be conceived as a challenge to any of the first three premises. It can be seen as a challenge to the first premise, by pointing out that the entities we call standing dispositional beliefs are actually most likely tacit anyway, thus denying the KD1 on its objectionable account of standing dispositional beliefs. Or it can be conceived as a challenge to KD2, granting the distinction between tacit and dispositional belief, but arguing any such distinction will not be robust enough to underwrite a kind distinction, since most instances of belief are tacit. Or it can be used to challenge KD3, making the case that the word "belief" is perhaps best affixed to tacit beliefs, since tacitness is the norm (though I do not much care for this last approach).

Whatever the best way to conceptualize its challenge to the argument, the resources on offer in the next two sections will amount to a serious challenge to this argument.

2.4 – Conclusion

I believe that we have not been given sufficient reason to think we have a substantive difference in kind that would warrant denying belief in the cases in section 1. However, full explication of this has yet to be given. But I have so far made the case that we have not been given any reason to think denying belief in such cases results in a superior explanation of any of the relevant phenomena. Instead, we have found again and again that a psychofunctionalist approach to the metaphysics of belief with the Fodorian aspirations of integrating folk psychology into the science of the mind, while satisfying the individuation conditions of mental states commonly held by philosophical theories of mental states, and common assumptions about the persistence conditions of mental states, just has no place for tacit beliefs. We have different theoretical starting points, not serious challenges to the claim that we have an infinite number of beliefs.

Section 3 – Belief Indicators

In section 1 of the paper, I made an initial case for the existence of an infinite number of beliefs. I started out by offering some cases where attribution of an infinite number of beliefs seemed intuitively compelling. I then entertained the possibility of denying belief in such cases, because they did not involve explicitly present, discrete, sentence or proposition individuable representations. I showed with additional cases that taking that route would result in some embarrassing belief denials. In section 2, I considered the arguments of the infinity skeptics and found them all wanting—ultimately concluding that they all boil down to assumptions about the very things that are at issue in disagreements about tacit beliefs and

the possibility of infinite beliefs. What is more they all seem to require denying belief in a number of cases wherein the norms of belief attribution seem to prescribe attributing belief.

What is left of the case against infinite beliefs then is the fact that some views about the nature of belief that are appealing cannot accommodate infinite beliefs and that tacit beliefs involve commitment to information that does not have a discrete, sentence or proposition individuable realizer in the mind.

I turn now to some further considerations that speak in favor of counting tacit beliefs as beliefs. The content of this and the following two sections will leave us in a position to see why the case for infinite beliefs is stronger than what is left of the case against them.

I will be using the term *belief indicator* throughout this section. So, I had better say what I mean by that. What I have in mind here are properties that seem to indicate that the person believes. These are the kinds of considerations we are apt to appeal to when trying to settle whether a person has a given belief. They can act as evidence to justify a belief attribution.

Some of the belief indicators I appeal to are used by theories of belief as full accounts of what it is to believe. But it is important to understand here that I am making no assumptions here about which, if any of these indicators may be constitutive of belief. I am not assuming any particular theory of belief to be correct here. Rather, it is the fact that a belief indicator is a good indicator of belief that tempts people to use one or more of them in their analysis of belief. It is the fact that in most cases of belief, many or all of the belief indicators I will talk about here are present that makes these indicators good candidates for an analysis of belief. And the fact that there are so many belief indicators is part and parcel of what makes settling belief attributions in weird cases so hard. It is what happens when some of the

indicators are present and others are not where we have to really get into the weeds with deciding what really matters for believing.

But as far as I am concerned here, there is no reason to get into those weeds. We can stay at the pre-theoretic level of belief attribution norms and simply recognize that any of the belief indicators below could be used outside of a philosophical debate on the nature of belief to justify a belief attribution. I will be in the business here of pointing out how pretty much the whole gamut of belief indicators seems to apply to tacit beliefs, like those outlined in section 1. The goal will be to show that tacit beliefs have far more in common with beliefs than differences. This will (a) explain why the norms of belief attribution do not prescribe belief denials in such cases, (b) provide resources for understanding what the consequences of such belief denials would amount to, and (c) show that our belief attribution practices are not tracking the presence of discrete sentence or proposition individuable mental entities. The indicators that we cotton onto to make judgments about what people believe are tracking cognitive commitments that do not require a realizer of that form. The conclusion of this section will make the case that the full import of these facts is something that has not been previously appreciated in debates about the nature of belief.

3.1 – Classical Belief Indicators

In order to get a handle on belief, philosophers have used metaphors like "belief is the map by which we steer" (Ramsey 1931). The idea seems to be that information we rely on to navigate the world is information that is believed. Tacit beliefs are freely brought to mind when their content is relevant and help us navigate the world. So, the map metaphor indicator seems to indicate that the tacit entities are beliefs.

But not all tacitly believed information will guide us in navigating the world. If we tacitly believe an infinite number of things, it is impossible for each to play a role in navigating the world. But this worry extends to beliefs that are typically thought to be realized by discrete sentence or proposition individuable entities as well. Not every belief ends up being important. All we can really say is that, should their content ever be relevant, one would rely on those beliefs to steer. So, too with tacit entities. So, the same counterfactual considerations apply to each. The tacit entities seem to be part of the map, even if there is plenty of map we will not ever need.

Another classic indicator is betting behavior.[22] The fact that one is willing to stake something on the truth of a proposition indicates belief in the proposition. And this property is generally present with tacit beliefs as well. I am willing to bet on any number x that is greater than zero being greater than zero. Steven would be willing to stake anything on the fact that he is not in Alpha Centauri. Chihiro is staking her time and the possibility of being locked out of her account on her belief about the required number. And Jenny is staking her reputation on the fact that she can successfully direct people to the location of her bathroom. (Can you imagine how embarrassed she would be if she failed!) So, betting behavior is present in all cases.

Another classic indicator is action.[23] How one acts is indicative of what one believes. How exactly such a property is to be understood need not concern us here. What matters is that even the most behaviorally demanding interpretations of the property will be present in tacit beliefs.

[22] One influential view of belief in this vein is *staking,* found in Ginet 2001.
[23] See chapter 3 for a fuller exploration of this issue and the literature surrounding it.

Another classical indicator is self-reporting. The fact that someone says they believe p and appears to be sincere is a good sign that they believe. It is not an infallible sign. But no one should doubt that self-reports of belief heavily correlate with what beliefs are actually held. And here too, tacit beliefs fare just as well. If asked whether one believes what one tacitly believes, one will unreservedly affirm belief.

Another classical indicator is phenomenology.[24] Beliefs seem to spring to mind as if our commitment to them was in some way antecedent to the present moment. We feel as if they were previously known or believed. And there does not seem to be any conscious act of inference that occurs.

Here too, tacit belief fares just fine. Numbers greater than zero pop into my mind freely in a way that does not seem to feel like an inference, but instead as something whose truth I have long known. Whether or not there has ever been a previous instance of me representing the number seems completely immaterial. It is highly doubtful that I have any of the sentences from the infinite series stored in memory, but they spring to mind will all the vividness and force of something that I had stored in my mind all along.

3.2 – Indicators from the functionalist theory of belief

There are many properties that have been proposed as constitutive of belief among functionalist theories of belief. And all of those properties extend to tacit belief.

It is commonly thought, for example, that one thing that is indicative of belief is that such representations be *poised* to guide action.[25] This suggests that they spring to mind

[24] How important phenomenology is to belief is a matter of debate. See Chalmers 1994 and Smithies (forthcoming) for views that takes its role as central.

[25] See Stoljar (forthcoming) for an exploration of poise and its role in the theory of mental states.

freely and are, in a sense, always ready to go. How such claims of poise are supposed to be understood in relation to problems of information retrieval need not concern us here. All that need concern us is the fact that tacit beliefs can spring to mind with the same ease as anything we would be tempted to say is non-tacit. If asked to find a number greater than zero, my beliefs about numbers greater than zero will produce as many numbers as you'd like. If Steven is asked to tell you where he is not, he will never run out of places to tell you (though he may soon run out of patience). And no matter where she is in her apartment, Jenny will be able to direct you to her bathroom with a pointed finger. Her relation to her bathroom is always known, regardless of where she is standing or whether she can articulate an English sentence to codify the relation between her present location and the bathroom. So, such information seems always poised to guide action, every bit as much as information that is stored in discrete, sentence or proposition individuable representations in memory stores.

Another common property to appeal to is the fact that beliefs are freely used as premises in inference to form new beliefs.[26] And tacit beliefs meet this standard as well. They are poised to be occurently represented and figure into inferential processes just as easily as non-tacit beliefs.

Another proposed property is evidence sensitivity or vulnerability.[27] While the infinite series does not seem to be sensitive to evidence, it is not sensitive to evidence in the same way that other a priori beliefs are not sensitive to evidence.[28] The rest of the tacit beliefs are

[26] The point is often attributed to Stich 1978, but it is not so clear that he makes the point. But this is one of the many ideas in orbit around the over-taxed term "inferential promiscuity."

[27] Van Leeuwen 2014 is one of many that take some kind of evidential property as constitutive of belief.

[28] We could also generate an infinite series that gets around this problem. If for example, I believe there are an infinite number of stars in the universe, I believe that for every number x, x is less than the number of stars. But suppose the latest cosmological models should indicate the universe is finite. I would give up an infinite number of those beliefs. Thanks to Dan Korman for this point.

sensitive to evidence in much the same way as regular beliefs are. If Steven found evidence that he was not in his bathroom, that he was in fact in a bathroom in Alpha Centauri, he would revise his beliefs about where he is and where he is not. If Chihiro saw that the numbers she was repeating did not match the numbers in the authenticator app, she would revise them, resulting in a new tacit commitment about what the required numbers are. If Jenny realized that she was mixing her new apartment up with her old one and misdirecting people to the bathroom location, she would update her mental map, and thus any sentence or proposition individuable tacit commitments.

Another proposed property is cognitive governance (Van Leeuwen 2014). The idea with cognitive governance is that beliefs interact with mental states of other kinds in an asymmetric pattern. If a belief and an imagining are used to draw an inference, the resulting mental state will be a new imagining, not a belief. If I believe "if p, then q" and imagine "not q", I will not believe "not p", I will imagine "not p." Tacit beliefs cognitively govern in much the same way. If I tacitly believe that 700 is greater than zero, and imagine there are 700 hobbits on the march, I thereby imagine a number greater than zero of hobbits on the march.

Another proposed property is *revisability* (Helton 2020). In order to believe p, it is necessary that one *can* change the belief. This can be thought of as a weaker version of evidence sensitivity. This one does not require that one does change their belief in light of evidence, but rather that it can be changed. If one is unable to change it, then it is a mental state of a different kind. Here too, tacit beliefs fare no worse. It is not clear how some of our a priori beliefs can be changed. For example, it is not clear how one could ever give up the belief that one is equal to one. Grasping the content seems to suffice for belief. And it is not

clear how beliefs in the infinite series could meet this condition, as it seems to be in a similar boat. But the challenges tacit beliefs face here are no greater than the challenges faced by the content of any such belief were it to be non-tacit.

3.3 – Teleofunctionalist indicators

Teleofunctionalist indicators are indicators that understand belief in terms of what it is aimed at. The most famous notion of this is that belief aims at the truth. The evidence for such a claim is transparency.[29] Questions of whether to believe p always seem to give way to questions of whether p is the case. Reasons to believe p are reasons that bear on whether p is true.

Other teleofunctionalist theories may appeal to evolutionary psychology, suggesting that belief aims at the avoidance of costly errors, at survival, or at something in that neighborhood (McKay and Dennett 2006).

Whatever the best teleofunctionalist theory may be, it would seem that there is no difference between tacit beliefs and non-tacit ones in this regard. Whatever aims of belief guide the formation and revision of beliefs will guide the formation and revision of tacit beliefs, because tacit beliefs are produced via the same cognitive processes and are explained by non-tacit information that is represented. So, any teleofunctionalist account of belief should imply the presence of the very same properties with tacit beliefs.

[29] The idea is typically attributed to Williams 1971, but has been fleshed out in, Velleman 2000, and Shah and Velleman 2005. The criterion is also used as part of Gendler's case for aliefs (Gendler 2008).

3.4 – Dispositional theories of belief

Dispositional theories of belief are the ones that are perhaps most sympathetic to the claim that we have an infinite number of beliefs, by their very nature, since acceptance of tacit beliefs is a common motivation for accepting dispositionalism. Dispositionalist theories of belief are non-representationalist. This does not mean that such theories require rejecting representationalism about the mind. Rather, it means that they reject representationalist analyses of belief (and presumably of other mental state types). Such theories are avowedly *superficialist* when it comes to belief. They take belief attribution to not be about giving an account of the cognitive at all. Rather, such theorists take the business of belief attribution to be more about the important evolutionary or social functions belief attribution plays. Their theories of belief then are not trying to integrate folk psychology into cognitive science as an account of the cognitive, but rather as a surface level practice that should be vindicated to some degree by an account of the cognitive, but not reduced to the cognitive.

Perhaps the most influential theory in this vein is Daniel Dennett's.[30] Dennett thinks that attribution of mental states is in the business of explaining and predicting behavior. Our attribution practices result in a high degree of determinacy and objectivity in settling questions about what is believed, but belief attribution, so conceived, should not be taken as an account of a realm of entities existing at the software level that match the individuation conditions and persistence conditions entities like belief are taken to have in our folk psychology. Our folk psychological attribution practices offer an effective theory for doing

[30] Dennett 1987 offers the best formulation of the view. Dennett 1981 offers additional insights into the application of the intentional stance. Dennett 2022 settles some longstanding questions about the metaphysics of the view.

explaining and predicting behavior, thereby making behavioral prediction a computationally tractable problem. But they are not in the business of carving the mind at its joins.

Thus by adopting this *intentional stance,* attributing possession of folk psychological kinds to agents, we can accomplish a great number of important things, but it would be a mistake to think that we are describing a realm of real entities when we attribute beliefs and desires. We attribute the beliefs, desires, and other folk-psychological states that an ideally rational agent would have in order to explain and predict behavior. The patterns we are picking up on are objectively real, and the results we get are often reliable, but this story need not in any way reflect the actual way the cognitive level works. It needs to suit our purposes of interpreting and predicting behavior.

Dennett's view then takes the explainability and predictability of behavior by attribution as the indicators of belief, which are constitutive of belief (or at least to be what act as truthmakers for belief reports).

And here too, tacit belief seems best understood as being as real as any belief. One's behavior is more successfully explained and predicted by attribution of tacit beliefs than by denying them. Consider, for example, if Eric is on a game show. We know Eric is going to be faced with the choice of opening a door for a prize. One door contains something worth zero dollars. The other door contains something worth 1,783 dollars. We do not need to actually think there is a discrete representation in Eric's head of the form "1,783 is greater than zero," to attribute to him the belief that 1,783 is greater than zero and thus predict what he will do. Nor do we need to wait until the number has made its first appearance in his mind to make the attribution. We can expect Eric knows which number is greater, whether he has a discrete representation with the relevant content prior to the event or not.

Another vein of dispositionalism follows on Wittgenstein (1958) and Ryle (1949), counting belief to just be a matter of having certain dispositions.[31] For Wittgenstein and Ryle, it was mostly dispositions to act that mattered. But a contemporary revival of the this line of thinking in Eric Schwitzgebel's work counts the dispositions that matter to be not just dispositions to act, but also to reason, and feel in certain ways (Schwitzgebel 2002). The view is close to Dennett's, but counts the truth of belief attribution as not being relative to an ideally rational agent, but instead to common stereotypes about belief. So conceived, whether a person believes p is a matter not of what is in their head, but rather whether they are disposed to act, reason, and feel in a way that we count as stereotypical of one who believes p.

Schwitzgebel's view is then constructed out of the indicator of steretotypes about how people behave, think, and feel when they believe.

Here too, tacit beliefs seem to fare just as well as non-tacit beliefs. People behave, think, feel, and reason in ways that are stereotypical of their tacit commitments just as well as they do the non-tacit ones. So, here too there is no difference.

That tacit beliefs fare so well on dispositionalist accounts is of course no surprise, as appeal to tacit beliefs is often a motivator for such positions.

[31] Dennett's view also takes a number of cues from this approach. Scwhitzgebel and Dennett's views end up being hardly distinguishable. But Dennett's view does not explicitly adopt any reductive language that suggests reducing belief to possession of dispositions. Instead, it seems as though dispositions to behave explain the success of the intentional stance and can be thought of as one possible explainer of the patterns Denett takes our attribution practices to be glomming onto. The classification of Dennett as a dispositionalist then is not uncontroversial. But with Dennett and Schwitzgebel's views being so close, a taxonomy that does not group them together is apt to be turning on distinctions that hide rather than illuminate similarities. See King and Zimmerman (forthcoming) for a taxonomy of philosophical work on belief that makes this grouping.

3.4 – Conclusion

I have walked through a number of belief indicators in this section, in order to show that tacit beliefs have all the hallmark properties of belief. They explain and predict people's behavior, reasoning, and feeling, when their content is relevant just as well and as often as beliefs we are more apt to call non-tacit. They are used to navigate the world, figure into inferences freely, cognitively govern mental states of other types, meet standards of evidence sensitivity and revisability that are common to belief, and influence betting/staking behavior.

Of course, a skeptic may wish to say that all of this commonality is explained by the fact that such states are dispositions *to believe*, and that when the relevant representations are realized, they function just like beliefs, because they then become beliefs. So, it is not as if the skeptic has to run away scared at this list of properties.

But what I do hope this shows is that tacit beliefs seem to fare just as well when our belief indicators are applied, that the norms of belief attribution seem to recommend attributing such beliefs to agents and expecting them to act according to those beliefs; that denying belief will do a worse job of predicting behavior than attributing will; that the status of a representation as having previously been tacit or having had explicit representation prior to it springing to mind is conceivably not knowable via introspection; and that the norms of belief attribution are not tracking the presence of discrete, sentence or proposition individuable representations.

If that last point is correct, we can also draw an unpleasant conclusion for the infinity skeptics on its basis. If none of our belief indicators track the distinction between tacit and non-tacit beliefs, then the evidence for belief that they provide is of equal strength for both

tacit beliefs and the ones claimed to be non-tacit. And any attempt to deny that tacit beliefs really are beliefs will amount to a claim that the evidence that such indicators provide is not very good evidence, since that would amount to them getting it wrong more often than they get it right (since the quantity of tacit beliefs far outstrips the non-tacit ones). Thus, any denial of the claim that tacit beliefs are really beliefs amounts to a denial of the validity of our belief indicators, and it is hard to see how that denial would not extend to any theory that utilizes any of our belief indicators as evidence, to any theory that takes any of them to be constitutive of belief, to any theory that takes them to be an important part of the story of how we know about beliefs, and to any theory that seeks to vindicate our folk psychology, which attributes belief on the basis of such indicators.

Thus a denial of tacit beliefs and of infinite beliefs threatens to undermine any theory one may wish to save by making such a denial. Refusing to acknowledge such indicators as good evidence undermines the very edifice all work in the philosophy and cognitive science of belief is built on. But accepting such indicators as evidence means that we have equally good evidence for the fact that what is tacitly believed is believed just as much as what is explicitly represented. Our evidence that belief does not require the presence of a discrete, sentence or proposition individuable representation then is far greater than our evidence that belief does require such a thing, because a denial of this claim amounts to a denial of the validity of the evidence body that makes up the core of our study of belief.

Section 4 – Information Architecture

I have made the case that there is much that speaks in favor of infinite beliefs and what speaks against it seems only to be that certain theories of belief do not seem to be able to accommodate the claim that we have an infinite number of beliefs.

I now turn to implications. Supposing it is true that we have an infinite number of beliefs, and we start with this as a constraint on mental architecture, what does that look like? Can we make sense of ourselves as having an infinite number of beliefs, while holding onto a materialist theory of mind?

One obvious tactic for doing so is to embrace dispositionalism, thereby divorcing the theory of belief attribution from the theory of the cognitive and denying that beliefs are entities in the mind. While I recognize the motivations for such views as compelling, I'm not terribly fond of such views. And they have not won many converts.

I will instead operate here from the assumption that what we want is to make sense of beliefs as real features of the mind whose presence or absence does not dependent on the adoption of a stance (e.g. Dennett) or popular stereotypes governing belief attribution norms (e.g. Schwitzgebel). I will operate from the assumption that we want an account of mental architecture that can make sense of belief possession being stance-independent, but which can nonetheless accommodate an infinite number of beliefs.

Is such an architecture possible? I believe the answer is a resounding yes—that it is not only possible, but actual. I will here offer a picture of what such an architecture might look like, make the case that we have good reason to think our mental architecture is more like what is here described than in rival views, and unpack the implications of such an architecture for the theory of belief.

This section has two slogans it will attempt to vindicate.

Slogan #1: Storage and retrieval are modally transformative.

Slogan #2: Tacitness is the norm for storage and retrieval.

The meaning of and justification for these two slogans will, I hope, be clear by the end of this section.

4.1 – Intuitive Ideas about Mental Architecture that are Apt to be False

It is commonly thought that entities like beliefs have a life that looks something like this. There is a discrete belief formation event for that belief. The belief formed has sentence or proposition individuable content. The belief then migrates into long-term storage, retaining its form and content, where it can be freely recalled whenever its content is relevant. The belief stays in the storage either for the remainder of the agent's days or is forgotten.

At the core of this picture is the idea of a mental state as a discrete entity with well-defined persistence conditions that are tied to its discreteness, form, or content.

Philosophers have fleshed this core out in a number of ways. In the classical Fodorian architecture, the entity that is formed in the belief formation event is a mentalese sentence. The sentence is then sent to a "belief box" (Fodor 1987). There are boxes for other mental state types as well. And information storage and retrieval is something like sentence migration. The sentence moves into the belief box to be stored. It is either pulled out or copied when it is needed for occurrent mental processes.

Fodor's later view was instead that beliefs involve abstract entities, called *propositions* as well (Fodor 1994). But the overall picture was much the same. The mentalese sentence had as its content a proposition, and believing was a matter of having that mode of

representation with that content in the belief box or some architectural equivalent for the full duration that one believes the relevant proposition.

On other views, the storage may be weblike, or like unto many webs. And on other views, the storage may need only involve some kind of tagging system to keep track of which representations are belief. All such views hold onto the presumption that believing is tied to the persistence of some discrete entity, and the persistence of its form and content.

Such views seem to have no place for tacit beliefs, either having to deny they are genuinely beliefs, as Audi does, or having to handwave away such entities with language like, *the primary case of a belief involves a representation* (e.g. Quilty-Dunn and Mandelbaum 2018)—whatever that means.

4.2 – Optimizing information architecture

But there seems to be little reason to suppose that the sentence migration picture of storage and retrieval is optimal information architecture. We cannot have before our mind constantly all of the beliefs we have in storage. Rather, information has to be found when its content is relevant, if it is to guide action or service as a premise in reasoning. And storing information as an array of sentences seems inefficient for a number of reasons. I believe such reasons will become most apparent if we consider an information storage task.

Consider for example if we have an array of thousands of four digit numbers that we want to store. One way we could do so is by just putting them in the order we received them. Sometimes it is helpful to store information this way. But only when the conditions for retrieving information are going to mirror the conditions we learned them in. Consider for example learning to count to ten in a language you do not speak as a party trick. You may be

able to produce all ten numbers as often as anyone asks, but when asked how to say "seven" in that language, you may have to count on your fingers for each number, starting from the beginning, until you get to the seventh number. You thus have the information stored in a way that allows you to produce each number reliably, from one to ten, but it can only be retrieved by going through the information ordered in the way you have learned and rehearsed it. Extracting the information without going through that ordering is difficult or perhaps even impossible.

But suppose we do not want to have to access our information in the order it was processed. Then it seems as though we need to find a way to organize it, which will reflect the expected conditions of retrieval in a useful way.

Let us suppose that our expected conditions of retrieval are that we will want to be able to take any new number given to us, and determine whether the same number is on the list of numbers we are storing.

Our constraints then are that we need the numbers to be searchable in a way that allows us to compare it to the number we are given, we need to be able to quickly decide what numbers we can ignore and which ones we need to look at, and it would be great if we could keep the amount of information stored as small as possible. Less stored information means less information that has to be combed through.

Interestingly enough, we do not actually have to store all of the numbers to do this and in fact we are likely to be better optimized to retrieve information quickly, by minimizing the amount of information explicitly stored and maximizing tacitness—so much so that we do not need a single number in our original list to have an explicit representation in the mode it was given.

Here is one way we could organize our body of information to reflect the structure of our intended use. We could group all the numbers that start with 1 together, all of the numbers that start with 2 together, and so on. We could then mark each grouping and delete the first digit off of every number, in order to save almost 25% of our storage costs, thereby making it so that we have do not have to comb through the first digit of every number. Rather, if the number we are given starts with 4, we move to the grouping of numbers starting with four, in order to start our search there. This process of grouping and deleting a digit can be repeated with the second digit, if doing so speeds up the process. Then, instead of thousands of four digit numbers, we have to store 81 groupings of two digit numbers or 9 groupings of three digit numbers. We find what we are looking for by going to the relevant grouping. And the fact that what we are looking for does not have the form of what we were asked to store is not only not a problem, it is part and parcel of what has improved our storage and retrieval process.

We could further improve our optimization by going into each grouping, and ordering the numbers there sequentially. For every grouping of consecutive numbers, we could store the first and last of the grouping, and mark that everything in between those numbers is also in the array, while deleting their explicit representation (thereby making their representation completely tacit). We could also delete every zero at the end of every number that ends in zero. Since we know that all the digits have to be four digits long and we know that we only deleted zeros, recovering the deleted zeros via a rule should be easy enough.

We have reduced the size of the information we have to store significantly and made the information more searchable. We destroyed its mode of presentation in a systematic way, in order to optimize our body of information for storage and retrieval—doing so in a way that

reflected our expected retrieval conditions. Rather than preserving the mode of representation for each number, we destroyed them in a way that made it easy to find them. We can use rules to recover their original form, if needed. But we can put them into different forms just as easily. Or we can use rules to settle questions we have about what numbers originally went in without *ever* recovering their original form. We have thus made it so that all of our numbers tacitly represented, in one of two ways. Some were tacitly represented by deleting information, and replacing the deleted information with a grouping, thereby eliminating redundancies and chunks of stored text. Other numbers were made tacit by being deleted entirely and replaced by some marker for consecutive sequences of numbers. We thereby significantly reduced the amount of numbers we have to look at in order to settle questions about what numbers were in our original list of numbers.

It would seem then that principles of good information storage do not require preserving the modes of representation of inputs to the system. It would seem also that retrieval processes can transform mode and content. (By transforming content, I do not mean to say that it can retrieve content that is not stored in some way, but rather that what can be retrieved need not be an exact match for the content of anything that went in. In other words, representations can come out that never went in, but no information can come out that is not explained by what went in. More on this point to come.)

Information storage and retrieval is about organizing information in such a way as to ensure that it is reliably retrieved when needed. This means reducing size, by destroying modes of representation in systematic ways, organizing the information in such a way as to be optimal for our expected retrieval conditions, and maximizing tacitness, in order to

reduce the amount of information that has to be combed through to find necessary information.

Now, our toy example was using numbers, but I think it gives us a good idea of why the classical Fodorian picture is going to be implausible. First, it is a sentence first approach to mental architecture. But the structure of sentences and lists of sentences does not seem to reflect an ordering of information that is optimized for storage and retrieval. We cannot order sentences alphabetically and then search through the whole array each time we need information. Information needs to be ordered in such a way as to make finding information easy. And preserving modes of representation seems to be completely unnecessary to do so, and may even be detrimental for doing so.

Consider, for example, these four sentences:

(S1) James is tall.

(S2) James is kind.

(S3) James is ugly.

(S4) James is tall, kind, and ugly.

How is a sentence-first mental architecture to make sense of a person who believes some or all of these things? Does the agent perhaps believe S1-S3 and then infer S4 as needed? Or perhaps, the agent believes S4, and then infers S1-S3 as needed, from S4. In the former case, S4 is not believed, and in the latter case, S1-S3 are not believed. Only in the case where all four are present as discrete representations does the sentence-first architect get to affirm belief in all four. But a well optimized storage and retrieval process could reliably produce all four as output, regardless of whether an inference was made moving from one grouping to the other.

Consider an alternate architecture, wherein there is just a mental file for James, with the properties of "tall," "kind," and "ugly" listed. Such an architecture allows S1-S4 to be produced as needed, but contains no sentences and requires no act of inference from S4 to the others or vice versa. Such an architecture also eliminates redundant "James" instances from the four sentences, and eliminates redundant uses of property words in different sentences.

Such a picture of mental architecture seems to be far more optimal. It reduces redundancies, reduces size of stored information, and is stored in a way that is more optimal for information retrieval, because it involves less information to be sorted through and organizes information in such a way as to mirror expected conditions of retrieval. If we are thinking of James, we pull up the mental file on James, and have the information ready to go, with the ability to put it into sentences if needed.

But this is not a method of information storage that vindicates many platitudes about belief that are offered by philosophers. This is a picture of information storage that destroys modes of representation and allows that the inputs and outputs of storage processes may differ markedly. "Dwayne Johnson is tall" and "Dwayne Johnson's performance in that movie were terrible" may go in, and "The Rock is tall and a bad actor may come out," without any conscious acts of inference being performed and without the sentences that went in ever seeing explicit representation ever again. This is a picture of storage wherein modes of representation are not preserved and completely new and novel representations are produced as remembered information, despite differing in form and content from any sentences that were input. It is a view of information architecture that indicates that beliefs typically do not have persistence conditions related to the presence of an explicit and

discrete, sentence or proposition individuable representation. It is one where storage and retrieval processes do not respect the boundaries of discrete propositions or sentences. Instead, information is transformed and combined in ways that may make the inputs and outputs of memory processes very different.

4.3 – Principles of Storage and Retrieval

The previous subsection has been aimed at pointing out how a view of mental architecture that counts sentences as the primary unit of storage and retrieval, and sentence migration as an account of storage and retrieval processes, is unlikely to be a well-optimized system of storage and retrieval.

I now turn to systematizing what we have learned. Below I articulate what strikes me as reasonable norms governing information storage and retrieval optimization. I do not claim that they do in fact operate in human minds—though I find the claim plausible. Rather, I claim them to be an account of what *good* information storage and retrieval looks like—or at least a start to telling that story. To what extent evolution has optimized our storage and retrieval processes is something that requires study. What import this has for the theories of belief and mental architecture will be unpacked in the next subsection.

We can formulate some of the principles governing information storage optimization as follows:

Shrinkage: information storage processes need to shrink the amount of information stored, in order to reduce the amount of information that needs to be combed through to locate and retrieve information.

Modal destruction: storage processes do not preserve modes of representation, but rather destroy it in a systematic way to ensure its recoverability by application of a rule. The function of such destruction is to optimize information for search and retrieval.

Retrieval dependent modality: information is destroyed in such a way as to optimize it for retrieval. Its mode of representation is dictated by constraints on search optimization and expected retrieval demands.

Retrieval dependent organization: Information is organized in a way that optimizes it for search and retrieval. Thus, how information is organized depends on retrieval expectations and the ability of these processes to tailor to retrieval expectations.

Tacit maximization: Anywhere that information can be made tacit without slowing down search and retrieval processes, that information will be made tacit. Less explicit representation means less information to be combed through, in order to locate relevant information.

We can also formulate some principles of information retrieval:

Modal construction: modes of representation used in occurrent mental processes are constructed as part of the retrieval process.

Computational Retrieval: retrieval processes are genuinely computational in their own right. Retrieval processes have to make inferences about what is tacitly represented and apply rules to construct information based on how information has been deconstructed. This means that recursive information structures can generate information that is the product of inference, which are delivered to conscious reasoning processes as information that is known and remembered. Thus, a belief with content p could have never had a discrete

formation event, could have never previously been represented before, and yet still be brought to mind as if it were known all along.

Accuracy trumps mode: retrieval processes aim to preserve commitments, not antecedent modes of representation. Thus, in cases like S1-S4 in the previous section, the mode of representation of the representations that generated commitment to the truth of those sentences can be different from the mode of representation that is produced by the retrieval process. Thus, it does not matter whether the agent learned S1-S3 or S4, S1-S4 can each be the product of memory retrieval processes, without any act of inference from S4 to the rest or vice versa. Such transitions can occur as part and parcel of decomposing and reconstructing information, without any act of inference from one grouping to the other. Thus, not only can the mode of representations change, beliefs whose content have never antecedently been represented (which could only be reached by drawing inferences, were one's starting point the beliefs that had explicit formation events), can be recalled freely despite such inferences never being drawn. The storage and recall processes aim to preserve the information, not an accurate list of previous representational modes.

4.4 – The implications of optimized storage and retrieval processes

I have just made the case that the norms governing information storage and retrieval do not mirror common assumptions about, nor are they likely vindicate common ideas about the nature of mental states or their persistence conditions. I of course cannot claim to know that our information architecture is optimized in the ways described above. But I do think the picture that emerges from such considerations actually makes good sense of how it is that we can have an infinite number of beliefs. It also makes good sense of the infinite

productivity of speech and why it is that we can almost never remember the exact words someone used, but can usually put what they said into our own words (a new mode of representation).

On such a picture, when it comes to stored information, tacitness is the norm, explicit representation is to be eliminated when possible, and explicit representation is typically modally transformative. The beliefs that are tacitly held are or at least can be infinite in number. But the quantity of representations that are used to produce occurrent representations of believed information is finite. Non-tacit representation explains all tacit commitments, but there is generally no way to know whether any given commitment is stored tacitly or explicitly, and modal transformations may make all beliefs tacit in some sense. It is only in weird cases, like those that have been used to defend tacit beliefs, that we can have a pretty good idea that information must be tacit. What is more, because it is likely that modes of representation are destroyed as part of the storage process, there is a sense in which any sentence or proposition individuable entity can only be tacit, since there may be no mode of representation corresponding to exactly the mode of representation being used to individuate the belief.

If such a picture of information architecture is even remotely on the right track, it has huge import for the theory of belief. First, it means that beliefs do not persist as discrete entities that can be identified with representations in the mind, since discrete representations do not persist through the storage and retrieval processes. Second, it means that any theory of belief individuation that individuates beliefs by appeal to sentences or modes of representation is likely to be false, since belief persists through changing modes of

representation.[32] Third, it means that any attempts to sweep away tacit belief is unlikely to be fruitful.[33] Fourth, it means that introspectively, we cannot tell what was tacitly or non-tacitly represented in our memory stores. Fifth, it means that there is no reason to deny we have an infinite number of beliefs, in order to integrate the theory of belief with cognitive science.[34] It is views that have no room for tacitness that are beginning to look incompatible with cognitive science.

4.5 – Independent reason for thinking that storage and retrieval is not sentence migration

In 4.1, I presented an intuitive picture of mental architecture and the persistence conditions of mental states. In 4.2, I offered some examples that showed why that would be bad mental architecture. In 4.3, I offered some principles of storage and retrieval optimization, which, whether or not they reflected the organization of the human mind, would amount to better information architecture. Then in 4.4, I explained what import this would have for the theory of belief, if such principles reflected the structure of the human mind.

I now turn to some independent reasons for thinking that the principles in 4.3 do in fact better mirror our cognitive architecture, than the intuitive picture on offer in 4.1 that is often presumed by a great deal of work in the philosophy of mind.

The first reason is that language is a late comer in our evolutionary history. And since sentence-first architecture amounts to poor optimization, there is no reason to think our minds would evolve in that direction. Our ancestors already had to have information storage

[32] E.g. Fodor 1978, Salmon 1986, and Fodor 1994
[33] E.g. Audi, Sellars, de Sousa, Powers, Richardson, etc.
[34] Audi was quite convinced in Audi 1994 that jettisoning infinite beliefs was necessary for the task.

and retrieval processes sufficient to even learn language long before language could enter the fray.

One may counter in Fodorian fashion that thought itself is language like, thereby granting language has been on the scene for as long as thought has been. Such a commitment may get around the first reason, but it is going to be powerless against the rest, so I will not explore possible challenges to it here.

The second consideration is that that sentence-first architectures are unlikely to have evolved in the first place. Such a poor information architecture would require significant computational abilities to ever be useful for storage and retrieval. So, information storage and retrieval processes that do not require such computational abilities are more likely to have gotten a head start in our evolutionary history. And since such architectures are more optimal anyway, there would be no reason for evolution to favor a process that is less efficient and requires significant computational abilities to ever be useful.

The third consideration comes from confabulation. Confabulation is most well known from cases of split brain patients,[35] wherein the visual field of the split brain patient is divided and instructions are given to the non-speaking brain-half. The speaking brain-half is then asked why they did what they did—the real reason relating to instructions given to the non-speaking half, but withheld from the speaking half. The speaking brain-half confabulates a reason for the behavior, without realizing it is a confabulation—as confident as they ever are when explaining their behavior.

If confabulations are indistinguishable from information that is delivered from information recall processes, this suggests it is the product of information recall processes.

[35] See Hirstein 2004 for exploration.

And if information recall processes are genuinely computational and constructive, then they are producing information that is being constructed out of whatever is found by search processes. Thus, we might explain confabulation like this: search processes found information that seemed relevant, constructive and computational retrieval processes produced a deliverance that had the highest probability of being the explanation for why the confabulator acted, and this deliverance was not distinguishable from any other deliverance of memory retrieval processes, which are produced in exactly the same way. This could explain why the confabulator is so confident.

If memory retrieval processes were not constructive or capable of inference, then the confabulation would not be the product of memory retrieval processes. So, it would be mysterious as to why the confabulator is so confident in the information. One may wish to point out here Gazzaniga's explanation as this all being the product of an interpreter module, which interprets behavior to escape the possibility of it being the product of memory retrieval processes (Gazzaniga 1985). But this just pushes the problem back a step. If the confabulator is confident because the interpreter module was confident, why was the interpreter module confident? Clearly, the deliverances of the interpreter module, if it exists, often distinguish between what is known to be the right answer and what is just a best guess. It is not unusual for us to not remember why we did what we did and not confabulate, or to half-remember and not be confident in our answer. So, it is not as if the interpreter module is not capable of making such distinctions. So, we need to explain why the interpreter module had no clue that this was a confabulation. The principles of information architecture I offered above would seem to answer. If search processes go on the hunt for information that is relevant, and construct and make inferences to deliver information to the interpreter

module, then the interpreter module likely did not recognize it as a confabulation, because it looked exactly like the deliverance of memory retrieval processes always do. The memory retrieval processes had no way of assessing whether what they found was right. And the interpreter module saw nothing about the information it received that was suspicious. The assumption that recall processes are constructive and computational and that the confidence of the confabultator is to be explained by the fact that the information was delivered by recall processes seems to be an illuminating way forward on confabulation.

A fourth consideration speaking in favor of the principles of information storage and retrieval outlined above is that we often cannot remember the mode of representation that generated a belief. If someone tells us something, and then asks us to explain what we just heard, we may say back something similar to what they said, but it is highly unlikely that we will be able to say it back word for word—in the exact mode of representation it was given. This suggests that the mode of representation was not preserved.

Further, if you think of what it is like to listen to a lecture or speech, you are not apt to recall much of the exact wording of the speech. And it is not as if you are making consciously making inferences about what is said the whole time. Despite not making such inferences, it is likely that you will be able to explain what was said in your own words—generating a series of representations that had never before been present in your mind. There will be new sentences. The boundary between sentences and propositions in the speaker's presentation will be different from the sentential and propoositional boundaries in the original presentation. And information produced will be information that could be inferred from the sentences contained in the speaker's presentation, but you will not be (or at least need not be) aware of ever having made such inferences. This again suggests that storage

processes do not preserve modes of representation, do not respect the boundaries of representations and propositions, and aim instead to preserve commitments that can be used to construct new modes of representation on the fly. And if that is right, then memory processes destroying and constructing informational modes as part and parcel of the storage and retrieval process is a likely explanation.

4.6 – Conclusion

I have made the case here that attending to constraints of information optimization suggests that information architecture is likely to undermine many common ideas about the persistence conditions and composition of mental entities, like beliefs. Rather than thinking of information retrieval processes as combing through an array of sentences like an old box of records, we should think that information storage is destructive in a way that reduces the amount of information that needs to be stored, changes modes of representation, and maximizes tacitness. And we should think of retrieval processes as constructive, generative, and inferential, in their own right. This does not mean that they are capable of drawing all inferences or of engaging in complicated reasoning. Recall processes are not a substitute for conscious "slow thinking." But we should think they are capable of making some basic inferences that can be used to construct modes of representation by application of rules, extracting tacit commitments, and that the transformations that occur in storage and retrieval can be functionally equivalent to inferences made on sentences.

This is a far cry from the boxology that predominates much of the philosophy of mind, wherein a belief is formed, migrates to a belief box, and is retrieved when needed. But if the picture of mental architecture I have outlined here is on the right track, it suggests that there

is no hope for such an architecture. But there is good sense to be made of infinite beliefs. And there is good reason to think that evolution would favor creatures that could maximize information responsiveness by encoding an infinite number of commitments in a finite architecture, which could be retrieved at will and at high speeds.

So conceived, the idea that we have infinite beliefs is not an embarrassment to be swept away, nor need it be a motivation for denying the reality of beliefs or their role in the mind, as the dispositionalists do. Instead, it is a massive clue as to the nature of our mental architecture that can help us understand what beliefs are and the role they play in the mind. Denying belief in the case of tacit beliefs saves a simplistic theory of mental architecture; it does not save belief. If what I have argued here is on the right track, tacitness is the norm. It is discrete, sentence or proposition individuable representations persisting throughout the life of a belief in a single mode of representation that is unusual.

Section 5 – Back to our original question

So, should we accept that we have an infinite number of beliefs? I believe so. I do not think that we have a slam dunk case for it, but I do think that quite a bit more speaks for it than against it. Further, I think that psychofunctionalists have no reason to fear the claim that we have an infinite number of beliefs or think that combined powers of cognitive science and reference magnetism will lead to an identification of belief with some discrete entities that persist for the entire time a proposition is believed. Nor need acknowledging tacit beliefs as genuinely beliefs lead us to superficialism[36] about belief. But it does mean we need to be more sophisticated in our approach to mental architecture, in our theories of

[36] E.g. Dennett and Schwitzgebel

mental states, and in our thinking about the persistence conditions of mental states. And rather than trying to subsume a theory of tacit beliefs to a theory of explicit beliefs, it may be time to go the other direction subsume all belief to a model of tacit belief, counting the production of retrieval optimized tacit commitments as a goal of storage architecture. The preservation of representational modes, like sentences, seems an odd goal for mental architecture.

Further the picture of mental architecture on offer here suggests a starting place for thinking about what the criterion of tacit belief actually is.[37] If memory retrieval processes are constructive and computational, then this suggests that what we tacitly believe is determined not by entailment, but instead by what inferences memory retrieval processes are capable of making and what transformations occur in storage and retrieval. If the deliverances of our memory retrieval processes require us to draw further inferences, even if they are simple, then that inferred information is not a deliverance of our memory retrieval processes, and thus something that was not antecedently believed. But if our retrieval processes had to engage in some recursive computation to produce numbers for representations of propositions in the infinite series or apply a rule to construct a computationally effective mode of representation for information that was stored (even if the output does not have the form or content of any representation previously realized in the system), that does not count against its deliverances having been antecedently believed. The limits of tacit belief would thus be circumscribed by the limits of memory retrieval processes, since the function of memory retrieval processes determine what information is ready to guide action, to serve as premises in inference, and to report as true.

[37] For an account of several proposed criteria for tacit beliefs and arguments as to why none of them work, see Lycan 1986.

Thus, the picture of mental architecture on offer here not only vindicates the claim that we have an infinite number of tacit beliefs, it suggests a way forward on questions about what we tacitly believe. Such productivity seems yet another mark in its favor.

II. Beliefs without Believers

The philosophical study of belief has two distinct, but enmeshed traditions. One is a *language-first* approach, led by philosophers of language engaged in the study of semantics and philosophers of mind interested in the problem of meaning in mind, which takes as its primary object of study belief reports—sentences of the form, *S believes that p.* Philosophers in this tradition will tell you that "belief is a relation" between an agent and an object of belief (e.g. Russell 1910, Fodor 1978, Salmon 1986, Buchanan 2013). The other tradition is a *metaphysics-first* approach, which is much less concerned with belief reports, and talks of beliefs as entities or states—usually something along the line of representations serving a functional role that is typical of belief—and aims to give a satisfying account of their role in mental architecture, thought, and the production of action (e.g. Armstrong 1968, Lewis 1972, Block 1978).

What is the relation between these two conceptions of belief—between the belief relation identified by one tradition and the entities or states talked about in the other? A common reading of relational approaches to the study of belief in the language-first tradition, like the propositional attitude analysis, is that such analyses amount to a claim not just about the semantics of belief reports but to a metaphysical claim that identifies mental states themselves with relations (Buchanan 2013). Such an account explains the relation between the belief relation and the entities or states talked about in the metaphysics-first tradition, but it is hard to make any sense of it. It is not obvious that relations are the kinds of entities that could play the role that mental entities are supposed to. Nor is it obvious that they can have the properties we take beliefs to have. What is more, not everyone within the language-first tradition would even consent to such a description of their view. Nathan

Salmon, for example, takes the word "belief" to be ambiguous between the belief relation and the cognitive states, taking each to require their own respective treatments within metaphysics (personal communication). Stalnaker's early work talked a great deal about belief as essentially involving a relation between an agent and object of thought (1976a, 1976b), but the trajectory of his thought eventually landed on taking relations to abstract entities to be more a matter of utility than of proper metaphysics (2008).[38] So, it is unlikely that there is a univocal reading of the statement "belief is a relation" that is apt to be endorsed by everyone in this tradition who has been willing to say in print that "belief is a relation." Most philosophers who say it have just not been very explicit about what they take the statement to commit them to and any unity in the tradition seems likely to be illusory.

So, I will not be interested in pinning down a particular view about how the language-first tradition takes the belief relation to fit into the nature of belief, but instead will ask what *should* we think the relation is between the belief relations in belief reports and beliefs, the entities or states that figure into the causation of action, reasoning, and interactions with other mental kinds? Can we, for example, take the instantiation of a belief relation to be a necessary condition on the existence of belief? Can we read the individuation conditions provided by our preferred account of the belief relation onto the individuation conditions of beliefs themselves? Is there a one-to-one correspondence between beliefs and belief relations? Is there any sense to be made of the claim that beliefs *just are* relations? Or is it the case perhaps that beliefs, the cognitive entities or states, serve as realizers for belief relations, but have different identity and persistence conditions than belief relations?

[38] Stalnaker 2008, p. 109-110: "There is, alas, only a brain in the head, no pieces of information, no contents…Propositions…are abstract objects that we use to represent certain capacities and dispositions of people, and certain kinds of relations between them."

I will make the case here that the belief relation has little to do with the nature of belief. Beliefs, the cognitive entities or states, which figure into thought and action, have different individuation, identity, and persistence conditions than belief relations—or so I will argue. My methodology for doing so will be to produce a number of thought experiments where it seems plausible to think that beliefs persist through the changing of belief relations and even through the expiration of any instance of a belief relation, thereby demonstrating the plausibility of thinking that beliefs and belief relations have different modal profiles. I will consider along the way the plausibility of an alternate position, which I call the relationalist position (representative of at least *some* work within the language-first tradition), which takes a belief relation to either itself be a belief or to be essential to the nature of a belief in some other way, and tries to account for the thought experiments by denying the persistence of beliefs through changes in belief relations. I will argue that while the relationalist can maintain consistency in accounting for the thought experiments, they do so at the cost of making beliefs a redundancy that is eliminable from the study of mind without loss. In other words, holding onto a relationalist analysis of belief comes at the cost of having to give up beliefs' role in cognitive science. I take the unattractiveness of such a cost to be sufficient reason for rejecting relationalist accounts of belief, like the propositional attitude analysis, arguing instead that we should distance the study of belief from the study of the belief relation and belief reports.

Section 1 will offer an account of the motivations for a relational analysis of belief. Section 2 will consider some entities that are commonly thought to be analogous to beliefs in important ways, which can be used to generate thought experiments for exploring the modal profile of beliefs and belief relations. Section 3 will offer a number of thought

experiments aimed at showing that beliefs can persist through changing belief relations and through the expiration of belief relations. While the relational analysis of belief can be consistently maintained to account for the thought experiments, I take it to be obvious by the end of the section that the costs of doing so are too great and that distancing beliefs qua entities or cognitive states from belief relations offers the more satisfying and defensible position. Section 4 offers some clarifications of my arguments, their place within the literature on belief, and the implications of distancing the study of beliefs from the study of belief reports.

1 – The Relational Analysis of Belief

Talk of belief being or essentially involving a relation goes back to at least Bertrand Russell (1910). Ambiguity about the metaphysical commitments of the oft-repeated sentence, "belief is a relation" go back at least as far.

The locus classicus for the analysis of belief reports in terms of belief relations and its importance for the study of the mind is Jerry Fodor's 1978 paper, "Propositional Attitudes." Fodor's case that belief is a relation is made by pointing out that "'believes' looks like a two-place relation." (1978, p. 502-503).[39] The case that it looks like one is made by looking at belief reports—sentences of the form *S believes that p.* Belief reports do look like they require that a relation hold in order to be true. They seem always to include a believer and a thing that is believed, which are related in belief. Producing evidence for such

[39] Fodor also appeals to Quine's Criterion of Ontological Commitment and to the claim that all alternatives seem hopeless, which I also cover in this section.

a relation is typically thought sufficient to license an inference to the conclusion that "belief is a relation."[40]

The fact that belief reports require a relation between an agent and an object of belief can be made intuitively, with a few simple observations. Take the belief report, "Sally believes that Biden is the president." It would seem that it involves a believer (Sally) and something the believer believes (that Biden is the president). It is hard to see how it could be true without the existence of these two things and their relation. Thus, we have a believer and a thing believed standing in a relation of belief. Thus, the analysis of belief reports as involving a belief relation seems like the best *face value* reading of such sentences.

We can sprinkle in a bit more rigor into our case by application of Quine's Criterion of Ontological Commitment (Quine 1953, Fodor 1978),[41] which requires performing an operation on sentences, which can serve as a test for their ontological commitments. The operation involves replacing everything in the sentence that can be so replaced with a variable, which is then existentially bound. Existential quantifiers in classical logic are traditionally thought to carry ontological commitment, so if the resulting sentences with existentially bound variables are in fact logical entailments of the sentences we started with, that will tell us what our ontological commitments are—or at least that is the Quinean story.

Take the following belief report:

(BR1) Elba believes that *Jurassic Park* is overrated.

Applying the criterion to BR1 requires replacing everything in BR1 that can be so replaced with a variable which is then bound by an existential quantifier. If the resulting

[40] The claim is also often taken to be so obvious as to be presumed without any argumentation at all.

[41] While the criterion's origin is in Quine's work, the line of thinking captured here is more reflective of Fodor's application of the criterion in Fodor 1978.

sentence is an entailment of BR1, then we have successfully prepped our sentences to apply Quine's Criterion. The two following sentences are potential candidates:

(BR2) Someone believes something,

(BR3) Someone stands in some relation to something.

It seems plausible that BR2 and BR3 are entailed by BR1. But whether this is so will depend on whether or not a face-value reading of the sentence is viable. Its syntax and logical form do not tell us whether the entailment holds. For example, take the following sentence:

(S1) Maddy is doing it for Tommy's sake.

S1 *looks* like it would entail a commitment to the existence of an entity called a *sake.* But there are no such entities (Dennett 1968). So, we cannot let the fact that BR1 *looks* like it will entail BR2 and BR3 convince us that it does. But if it does not, we should have some explanation of its semantics that can explain why. Many think there is no alternative reading of the semantics of belief reports that can make sense of them not carrying such entailments (Fodor 1978, Davidson 1991).

We can now apply Quine's Criterion. Existential quantifiers in classical logic carry existential commitment. So, if BR1 entails a sentence with existential quantifiers, then BR1 is committed to the existence of whatever is the value of a bound variable in the entailed sentences. If BR2 is an entailment of BR1, then it follows that BR1 carries a commitment to the existence of a believing agent and a thing believed. If BR3 is an entailment as well, then BR1 also carries with it commitment to the existence of a relation.[42]

[42] It is worth flagging that the case presented here has more to do with Fodor's application of Quine's criterion than Quine's own views. And even then, I am moving beyond Fodor's case in presenting BR3 as a possible entailment. Quantifying over relations is typically through to be a matter of second-order logic. Quine himself did not seem to think second-order logic even was logic (Quine 1970). Fodor's application of Quine's

Quine's Criterion has been the subject of much debate, and I will not have space to engage with such debates here. If Quine's Criterion proves indefensible, then there is still a strong intuitive case to be made for the semantics of belief reports requiring the existence of agents, objects of belief, and a belief relation, based on the facts that it seems like a good face value reading and that it seems difficult to see what an alternative reading could be. Thus, the face-value reading of belief reports, whether supported by intuition, application of Quine's criterion, or by the lack of attractive alternatives, offers compelling reason to think there is a belief relation, which is what is being reported in a belief report.

The case for the relational analysis of belief might also be made by pointing to the fact that it is common to individuate beliefs by reference to agents and contents—the relata of the belief relation. My belief that Biden is president and my belief that Harris is vice president can be distinguished by their differing content. But my belief that Biden is president and Sally's belief that Biden is president cannot be distinguished in this way. They can however be distinguished by the fact that one is my belief, and the other is Sally's.[43] Thus, it would seem that such relations can or do play a role in individuating beliefs, which might be taken as evidence for the relation of an agent and an object of belief being somehow integral the very nature of belief.

Within the study of belief qua relation there are a variety of accounts of the nature of the relation in question. The propositional attitude analysis of belief takes the objects of

---

criterion did not consider the possibility or need of quantifying over relations. But none of this matters for my line of explanation here, since BR3 is being offered as one thing that *might* be an entailment, rather than as actually being an entailment. And raising the possibility requires no stance on second-order logic or its relation to Quine's criterion.

[43] It is worth flagging once again that not everyone in this tradition will be on the same page about such things. Nathan Salmon, for example, would not want to individuate beliefs by relations, because he thinks that two people who believe the same proposition *p,* have numerically the same belief. He acknowledges that there is some sense to be made of talking in a different way, by appeal to the two different relations as two different beliefs, but he finds talking that way counterintuitive (Personal Communication).

belief to be propositions. Other candidates for the objects of belief include natural language sentence tokens in the minds of agents (Carnap 1947/1958, Quine 1956, Davidson 1967), natural language sentence types (Matthews 1994), utterances (Davidson 1989), interpreted logical forms (Harman 1972), mentalese sentence tokens (Fodor 1976, 1978, and 1987), or some other mental entity, likely a discrete, sentence or proposition individuable representation.[44] Other relational views may stick with propositions, but include a third relata, such as a mode of presentation (Fodor 1994).[45]

I will not attempt to assess here the strength of the case for the existence of this belief relation made in such work, but instead will simply take it as a premise that there is such a thing as a belief relation, which takes as its relata, at the very least, an agent and an object of belief—whatever the objects of belief may be. If it is instead assumed that there is no belief relation, then the case that such a relation has little to do with belief is already made. My aim here is to consider what such relations could have to do with belief, if they exist. So, taking the existence of belief relations as a premise makes sense for the current work, regardless of what the merits of the case for a belief relation may be.

2 – Useful Analogues

In order to think about the relationship between beliefs and the belief relation, I propose that we think of belief on analogy with other similar kinds, in order to generate cases to test our intuitions. Thinking about the persistence conditions of entities that are like beliefs in

[44] Many thinkers remain non-committal about the nature of content and the structure of representations, while still taking beliefs to be relations, which take as relata agents and representations and can be individuated in this way. So, any representationalist who has not made commitments about the nature of the representations involved, but is committed to such individuation conditions, fits this bill.

[45] Fodor 1994 commits to this view. The view is often attributed to Salmon 1986, but is view is actually that belief is a two-place relation, which we stand in by virtue of standing in another relation, which is a three-place relation of this form.

some important way can be a useful means for thinking about the modal profiles of beliefs and belief relations.

Functionalism about mental kinds is perhaps the most widely accepted approach to the metaphysics of mental states. While there is much dispute about what the function(s) of belief may be, the view that mental states can be understood in terms of function is widespread in philosophy and psychology. However, the functional science of mental kinds is still in its infancy. So, I suggest we take as an analogy a functional kind whose persistence conditions are much better understood: human organs.

Human organs, like human mental states, come into being in a way that seems to essentially involve a relation to human agents. But organs persist through changing hands and the extinguishment of their relations to humans they were grown in. A heart may be transplanted, while still being a heart. Its identity and kind membership are not compromised by changing hands. If Ashitaka dies and his heart is transplanted into Inoue, it is numerically the same heart, despite having changed hands. During the surgery, when the heart is removed from Ashitaka and before it is placed into Inoue, it is still a heart, despite bearing no relation to any agent. The fact that human organs are defined by their function, which relates them to humans, does not suffice to ensure their persistence conditions are tied to any particular relation to any human. It also seems conceivable that organs can be grown in a lab, having never had any such relation to any particular human. It is still a heart, even if it has not been transplanted into a human—its status as heart is the very reason it would be transplanted into a human. The organ's function and status as a human organ are all still matters involving relation to humans. But there is no particular relation to any particular human that is essential to the identity conditions or kind membership of organs.

A second source of useful analogy is software entities, like files on a computer or hard drive. Understanding the mind via a computer metaphor is common in philosophy and cognitive science. Computationalism, which takes thought to be computations performed on representations, is the most popular form of representationalism within cognitive science.[46] This approach often involved thinking of the mental as something like software and the brain as hardware.

If I have a hard drive with files on it, I can take it to different computers to have the file read/used. The file persists through changes in what hardware is accessing the file. And if the hard drive is connected to no computer at all, the file still exists despite not being functionally integrated with any computer. Thus, its integration with particular systems is not part of its identity conditions, nor does its lack of integration with any computer undermine its status as existing and being a software level entity.

If beliefs are like software entities like files, then the idea that they can change their relations to agents, like files change relations to computers, and that they could even persist despite having no relation to any agent, like a file on a hard drive in a closet could, would undermine the essentiality of any relation involving an agent as one of its relata.

So, if beliefs are or essentially involve relations with agents as one of their relata, then they seem to be very unlike software entities and they would seem to be exceptional functional kinds, having persistence conditions very unlike other better understood functional kinds. Their identity and persistence conditions would need to be tied to the fact

[46] The Representational Theory of Mind (RTM) and Computational Theory of Mind (CTM) are terms that are often used interchangeably. Among those who count them as distinct, CTM is thought of as a type of RTM theory (perhaps the most developed type), but not the only form an RTM can take. Thus, rejection of CTM is not tantamount to rejection of RTM, but rejection of RTM suffices for a rejection of CTM. See Pitt 2020 for an example of discussion of RTM and CTM in this vein.

that they are related to some particular agent in a particular way. We can use thought experiments to test our intuitions about whether we really do think of them as exceptional kinds, as the relational view would require, or whether we regard them as having identity conditions that are similar to these analogues.

3 – Thought Experiments Probing Beliefs' Persistence Conditions

I turn now to thought experiments that are meant to trouble the relationship between the belief relation and beliefs. Each case is meant to force our hand in clarifying the role of the belief relation in some way. Each case seems sensibly and profitably understood by taking belief relations to be non-essential to beliefs. Throughout, I will consider an alternative relationalist take on the cases that counts the belief relation as somehow being essential to the nature of belief.

I signal in advance that some of the thought experiments may be thought to be impossible, depending on one's preferred metaphysics of the mind. Nothing hangs on all of them actually being possible. I multiply cases here in hopes that pretty much everyone will find at least *some* cases possible. Anyone who is discouraged about the prospects of any single case is encouraged to just move onto the next one.

Before getting into the thought experiments, however, I have found that the readability of this section is improved by recapping and defining some terminology that is utilized herein. When I speak of *the belief relation*, I am speaking about the relation between an agent and an object of belief, which is supposed to be the relation discoverable by analysis of belief reports of the form, *S believes that p.* No other relations, no matter how relevant to belief they may be, will be referred to as *the belief relation.* When I speak of *the*

relational analysis of belief, I am speaking of the analysis of belief offered by those who say that "belief is a relation" and wish to say that this is somehow connected to the metaphysics of beliefs, the cognitive states or entities, either by identifying beliefs with the relations or taking the relations as being essential to the nature of beliefs. When I speak of a *relationalist,* I am talking about someone who endorses the relational analysis of belief. (I will take the relationalist to be my interlocutor throughout this section.) When I speak of an *agent relation*, I am talking about a relation that takes an agent as one of its relata. So, the belief relation is a type of agent relation.

3.1 – Conjoined Twins

Conjoined Twins Case: Mara and Quinn are conjoined twins who are conjoined at the head and share some small area of brain matter. Information stored in this area can be drawn upon freely by both Mara and Quinn. Representations in or essentially involving this area are functionally integrated in both Mara and Quinn's cognitive systems. They are drawn upon freely for use in inference and the guidance of action. No other information stored in the brains of Mara or Quinn is commonly accessible. Mara and Quinn's brains can be surgically dissociated, but doing so would require giving the shared area to one or the other, or destroying it. Surgically dissociating this area from one of the twin's brains would result in information stored in the shared area no longer being accessible to that twin. (Let us also assume that there is no redundant storage of the information in the shared area.)

Conjoined Twins Discussion: Conjoined Twins is a case that is plausibly understood as two agents standing in the same type of relation to one and the same mental entity. In

other words, it is a case of two distinct belief relations for every one representation in the shared area.

The case for there being two agents is made by the fact that there seem to be two brains (on the grounds that they are functionally distinct and surgically dissociable), two streams of consciousness, two different sets of memory stores (with only some small number of memory stores shared in common), and two different personalities. Whatever the conditions of agenthood or personhood may be, the distinctness of Mara and Quinn seems likely to fall out from the description of the case.

The case that it makes sense to talk about the representation being stored within this shared area is made by the fact that when connected to this shared brain area, both agents can produce representations that they would not be able to, were they surgically dissociated from this brain region. The idea that information can be in a certain area or at least a certain area can be essential to having that information can be supported by the mere fact that damaging brain areas can make information no longer accessible.

The case that there are beliefs in this shared area is made by the fact that there are representations stored therein that are playing the functional role of belief. Both parties are willing to report the truth of their content, use them freely as premises in reasoning, act upon them, plan with them, etc.

Prior to any surgical dissociation Mara and Quinn are both related to the same things. Within their shared storage is a representation with content p that satisfies the functional role requirements of belief, and it is integrated into both agents' cognitive systems. If either agent receives information that is inconsistent with p, then the belief will be discarded and may be replaced with a new shared representation that has the functional role of a belief.

If we take only a single representation with content p, which is functionally integrated into both cognitive systems and realizing the functional role of a belief for both parties, it seems to be the case that it suffices to relate both Mara and Quinn to the same abstract entity. So, on the propositional attitude analysis of belief, we have two belief relations.

If on the other hand the belief relation is a relation between an agent and a purely mental entity or between an agent, a mental entity (like a mode of representation), and a proposition, then we again seem to have two beliefs, because we have two agents, resulting in two distinct belief relations.

If we take a single representation within this area, there is some temptation to say here that there is only one belief that is shared between the two of them. There is only one mental entity, which has two distinct agent relations. If we adopt the analogy of organs, if Mara and Quinn shared a kidney or leg, we would not be tempted to double count the kidney or leg just because there are two agents using it. If two computers have access to the same file on the same hard drive, we would not be tempted to say that there are two files.[47]

While the strangeness of such a case may not compel our intuitions forcefully enough to move the intransigent in any particular way, it is worth noting here that it is at least not pre-theoretically obvious that the presence of multiple belief relations suffices for multiple beliefs here. And so, an assumption that we can help ourselves to identifying beliefs with belief relations seems to face a challenge here. The nature of the challenge is not that the relational view is decisively shown to be false, but rather its conceptual necessity

[47] When files are opened in a computer, they may be copied into short term memory systems. So, if the file is opened, it is possible that there are two copies of the file when it is opened. But when it is not opened, but just openable, there is only the one file.

and its status of being presumable without argument is undermined. We have prima facie reasons for thinking that beliefs' identity conditions may not be tied to their agent relations, and we can get a perfectly coherent account of beliefs not individuating according to belief relations.

The relationalist can of course make some moves here. She can say that we have one belief that is essentially related to two agents, allowing that the relation between belief and believer is not always one-to-one in the belief relation. Doing so undermines the case that we need agent relationality in order to individuate beliefs. If the belief relation that Sally bears to Biden being president does not preclude it also being my belief (since under this assumption, belief-object relations are not essentially one-to-one), then we cannot individuate our beliefs by reference to the fact that there are two different agents related to the same proposition.

The relationalist can also say that all that matters is that *some* agent is involved—not that there be any particular agent, thereby precisifying her account to tie the persistence conditions of belief to agent-relationality, but not to a relation with any particular agent. Doing so again undermines the case that we need agents to individuate beliefs, since the same belief could be related to different agents at the same time.

What is more, this route seems to abandon the relational account entirely, because it amounts to it being the case that there is no particular agent relation which the belief relation can be identified with, or which is essential to a belief's identity conditions.

Lastly, the relationalist can just say there are two beliefs here, because there are two agent relations. This seems like the simplest way forward for the relationalist—and the one the relationalist is most apt to take. But note that the case for a relationalist analysis of belief

is made by presuming belief reports of the form *S believes that p* to be indicative of the nature of belief. So, in insisting this be understood as two beliefs, by appealing to relationality of belief, she does not have any non-question-begging argument. She has decided in advance that she will consider only cases with clear one-to-one agent-object relations as her sole object of investigation (i.e. belief reports), analyzed belief as essentially having the properties of the only things she is open to investigating (belief reports), and is dismissing the possibility of one belief shared by Mara and Quinn on those grounds. So, she is dismissing the possibility on the grounds that she has a history of dismissing such possibilities, not because she has an argument that such possibilities should be dismissed.

She may try to defend her view by appealing to indexical beliefs, pointing to the fact that a shared representation, with essentially indexical content (Perry 1979), should be taken to suffice for two beliefs. Take for example, "I like video games" being stored within the shared area. When Mara accesses the representation ,"I" picks out Mara, and when Quinn accesses the representation, "I" picks out Quinn. Thus, two different relations to two different agents gives us two different beliefs.

There are a variety of answers that can be given to this argument, but many of them depend on substantive philosophical theses that cannot be argued for here. So, I will just list the many options.

One such route is to deny the existence of essentially indexical contents (Millikan 1990, Cappelan and Dever 2013). Another route is to deny that any such representation could be shared by the two, since it could not play the appropriate functional role of a belief. (E.g. if Mara likes video games and Quinn does not, then this could only be shared information if Quinn's access to the representation somehow indicated that Mara liked video

games, because she would otherwise not incorporate the representation into her cognitive system as a belief. Thus, representations of such forms would need alteration of form or some meta-belief about their contents to be integrable into both cognitive systems.)

Another route for pushing back is to point to the fact that this agent relation is not likely to be the belief relation anyway. The relation born to an agent here is a content determining relation (it determines the identity of the content), not the belief relation. So, while it is *a* relation between an agent and a content, it is not *the* relation that the belief relation is supposed to be. The belief relation is not a content-determining relation.

Another route is to just accept that in rare cases a relation to a different agent can suffice for a second belief, but deny its significance. If an alien organ X is not a heart, but X can be repurposed to serve as a human heart, because by cosmic coincidence X is structurally identical to a human heart, placing X into the role of a human heart can make it a human heart. Whether we wish to understand this as a matter of the organ changing kind membership or of it coming to have two kind memberships, the fact that different relations can sometimes suffice to change kind membership or identity does not show that the identity conditions of organs are tied to some particular relation to some particular agent. These are instances of changes in agent-relations merely co-instantiating with changes that do make a difference. What seems relevant here is the change in function. So, if the same representation plays a different function in each of their cognitive systems, that seems sufficient to grant that there are two different mental states. If I have a file that when opened by a mac is a picture of a dragon, but when opened by a windows computer is a picture of a kitten, the file realizes two different roles and thus we have two pictures realized by one file. But if it is the same picture of a dragon, no matter the computer that opens it, it seems

strange to suggest it is a new file, merely because of a new relation to a new pc. So, on this route, it seems as though we should say that additional belief relations are only sufficient grounds for the existence of an additional belief when the new relation is realized by production of a functional difference that suffices to make it a distinct entity. If the same representation is used by Mara and Quinn, but it has different content or for one of them it is a desire and the other it is a belief (e.g. I am the smarter twin), then we have two different mental states. If we had conjoined twins where an organ was somehow doing double duty, serving as a heart for one, but a liver for the other, the same lump of tissue would be functioning as two different organs. Thus, the mere possibility of getting two beliefs from one representational entity is of little significance for the question of whether or not non-indexical beliefs can be shared.

3.2 – The Brain Assimilator

Brain Assimilator Case: Goragon is an alien with a special power. He can take the brains of other sentient beings and connect them to his own, in order to assimilate the information contained in other brains. Doing so requires taking the entity's actual brain, smushing it against his, and generating connecting neural pathways to access the newly attached brain's neural pathways and thereby the representations stored in the brain and information about the functional role they were playing. When he assimilates another's brain, he does some revision of his beliefs for consistency (not every brain assimilated has beliefs that Goragon deems acceptable) but keeps all the acceptable beliefs from the assimilated brain, making no changes to their representational modes, nor to their storage locations in the original brain. He does not take on any of the other mental states of assimilated brains. The assimilated

brain does not retain the function of any person-level mental processes (it is not a separate consciousness) and whatever assimilated brain mass is not used for belief storage is restructured and repurposed, thereby destroying everything of the person whose brain it was except the saved beliefs. And let us also assume that there is no difference in functional role between mental states in Goragon's species and ours—they have all the same mental kinds with the same function.

Brain Assimilator Discussion: Do the beliefs of agents whose brains have been assimilated by Goragon persist through changing hands like this? I am tempted to say that they do and see nothing in the relationalist's case that speaks against doing so, besides the fact that they have decided from the inception of their inquiry to only look at typical belief reports, which are cases involving one agent-relation. I thus see nothing in their case for the existence of a belief relation that should sway anyone who is tempted to say the beliefs, the entities or states, have persisted.

The relationalist seems again to have only one viable option: to say that new belief relations have sufficed to generate new beliefs. However, it is hard to see what grounds there are for making such a claim beyond just the fact that she is in the habit of individuating beliefs by reference to agents, since that seems to be a tidy way of accounting for belief reports, and would prefer not to give it up. The physical brain tissue has persisted through this change. The information has persisted. The modes of presentation have persisted. The content (or relation(s) to it, if you prefer) has persisted. Their functional role has persisted. So, something very much like the beliefs that existed before has persisted through this change. It is hard to see any independent grounds for denying that the beliefs have persisted.

And it is hard to see why the semantics of ordinary belief reports would be telling sources of data on this issue.

3.3 – Brain Bisection

Brain Bisection Case: Joe undergoes a brain bisection. During the procedure, his body is irreparably destroyed. Fortunately, the doctors had two bodies that were perfectly good nearby. Being unsure which Joe would want, they decided to plant his left brain-half in one body and his right brain-half in the other, so that Joe could experience both bodies and decide which he liked. They install devices in each body that mimic the function of brain areas that are not symmetrical, thereby realizing any lost function (such as regulation of bodily processes, like appetite or heart rate). These devices do not contain copies of any folk-psychological states from the other brain half. When the two brain halves come to and have the situation explained to them, they each decide to keep their own bodies and live separate lives.

Brain Bisection Discussion: At some point in this procedure, most people will want to say that we got two people. I am inclined to think it occurs at the moment of brain bisection, but most people I have talked to about this seem to think that it happens when the brains are separated into new bodies. (Few people wish to say that Joe is now bilocated in two bodies.) Whatever the best way of understanding how we ended up with two people may be, and whatever their relation to the pre-bisected individual may be, we have at least one new agent, who has Joe's old brain matter and the information stored therein.

The only way the relationalist can hold onto her view in this case is by sticking to her guns and saying that new agents suffice to give us new beliefs, because they give us new

belief relations—that we need to add to our theory of belief formation that beliefs can generated via the surgeon's scalpel. I again see no non-question-begging reasons for going this last route.

3.4 – The Dead Man Cases

There are two cases here, in the interest of sidestepping some metaphysical worries that are beside the point.

Dead Man Case: Amir is recently deceased. But all of the beliefs he held can still be recovered by scanning his brain. Doing so reveals the structure of the brain, how information is encoded in the brain, what functional role representations have when the brain is online, and what their content is. Molly wants to know the location of Amir's buried treasure. She scans his brain and finds the information.

Dead Man with a Twist Case: Amir is recently deceased. But Amir had a cybernetic brain implant that he used to store information securely. Beliefs that contained sensitive information, like the location of his buried treasure, were stored securely in the device. The device ensured that such information could be accessed like any other beliefs stored in his head and that any representations of this information encoded anywhere in his brain (or other representations that might make the information recoverable by inference) were removed, with the exception of occurrent mental processes and states, wherein the information was being used. When the device was locked, Amir could not access the info located therein, making him unable to reveal the information in the face of torture. When unlocked, the information therein could be accessed freely. Removal of the device would have made it so the information stored on the device was no longer accessible to Amir at all

while alive, but so long as it was installed and unlocked, Amir could utilize the information freely when he was alive. Information stored on the device came to mind in a manner that was phenomenally indistinct from any other belief, and figured into inference and action guidance in the same way. The device shields the information from being recoverable by brain scans. Molly hacks into the device after Amir is dead and finds the location of the treasure.

Dead Man discussion: The dead man cases are cases wherein information that was stored in Amir's head is recovered after the agent has ceased to exist, and its status as information that was taken as true and that would guide action and inference, were such things possible, is recovered.[48] The question then is should we think that the mental states have persisted past the expiration date of their agent relations? It seems tempting to say that Molly has accessed beliefs—that these entities or states are the very same ones that figured into thought and action while Amir was alive. It seems tempting to say that if Amir could be brought back to life, before significant neural decay, that the beliefs will have persisted through the whole process.

The relationalist has to either deny that the representations in question are beliefs or opt for saying that the agent relation has persisted through the expiration of the agent.

If she opts to deny that the representations in question are beliefs, she must allow that something very much like a belief has persisted through the expiration of the agent relation. A representation with content and an identifiable function still exists. If we adopt

[48] The reason for having two versions of the case is in the interest of bypassing some metaphysical worries. Some views of belief may entail the impossibility of one of the two cases. So, in including them both, I can maximize the number of people who find some version of the thought experiment acceptable. If, for example, an externalist about content denies that examination of the form of a representation could reveal its content, she can imagine that the device in the twist case would spit out natural language sentences, whose content could be determined with a degree of accuracy comparable to any other form of linguistic communication.

the analogy of an organ, the heart of a dead man is still a heart, even though it stopped beating. If we adopt a computer analogy, the files on a dead computer's hard drive are still the same file types they always were, and can likely be recovered for some period of time after the computer has died. So, what could motivate the exceptionality of beliefs here as a kind that cannot persist through this change? There seem to be no independent grounds for claiming they cannot. The relationalist can only appeal to the fact that they are modeling the metaphysics of belief on the model of a belief report, an approach which assumes that there are no cases of belief without believers—that cases of reporting *someone's* beliefs are the only cases of beliefs existing that there are.

If she opts instead to say that we still have a belief relation, despite the believer being dead, it is hard to make sense of what is going on. She may point to the fact that we may still be tempted to refer to the belief as Amir's belief as evidence. But that does not seem to show anything. We often speak of things that belonged to dead people as if they still belonged to the dead person, but we do not actually think they are owned by the dead person. Speaking of Graybeard's map, after Graybeard's death, indicates that Graybeard may have once owned or created the map, without indicating that Graybeard currently is the owner of the map. Such semantic evidence is ambiguous between a reading that takes it to denote a current ownership relation and one that denotes a relation to a former owner—which is a relationship distinct from the current owner relation.

Perhaps she may want to think of belief as something like a time traveling relation and say that the belief still exists because of the prior existence of an agent and that it is this relation to a past agent that is the belief relation. Talk of the grave of a person, seems to work like this. Talk of Martha's grave is talk of the grave of a person that does not now

exist, but once did. But it is hard to imagine that the belief relation could work like this. Consider, for example, a case of belief change. If I stop believing p, and instead believe not p, it does not make sense to say that I still believe p, because at one time I represented p and it had the functional role of a belief, despite not having it now. My former belief that p no longer exists. So, it does not make much sense to allow that belief relations can exist at a time t, without both relata being present and bearing the appropriate relation at t. Otherwise, we would be able to report former beliefs as current beliefs, which we most definitely cannot.

So, it seems as though she does not have a plausible story to tell here about how the belief could continue to exist past the expiration of the agent-relation. Her best move seems to be to deny altogether that the belief has persisted. But doing so requires granting that something very similar to a belief has persisted—call it a schmelief. Schmeliefs it would seem are cognitive entities or states at the software level, with identifiable functional roles and content. But they are not beliefs, on this line of thought, because they lack any connection to a belief relation, like the ones found in belief reports.

While the relationalist can make this move, it has no advantages beyond rescuing the relational notion of belief from inconsistency. What is more, it raises a serious question about why we need beliefs at all in our theory of mind, since it would seem schmeliefs would do just fine. Schmeliefs are capable of playing any role that beliefs do, since they are just beliefs minus the belief relation. Schmeliefs can make sense of belief reports just fine, since there is no reason why schmeliefs cannot co-instantiate with belief relations. The positing of schmeliefs thus seems to make beliefs completely excisable from the study of the mind without loss. Thus, this move of saying belief has expired when the belief relation has

expired seems to be rescuing the relationalist analysis of belief at the cost of giving up on belief. If we think beliefs are important for the study of the mind, that is a fact that makes much more sense when we take belief to not be identifiable with or essentially involve belief relations.

Section 4 – Belief without the Belief Relation

I have just offered a number of thought experiments that I take to offer a compelling case that our beliefs can come apart from belief relations in important ways. The idea that the story we get from study of belief reports does not reflect the nature of beliefs is not new to the study of the mind. Matthews, for example, has argued that talk of belief, by appeal to a relation between an agent and an abstract entity, is something more akin to appealing to an abstract entity in measurement. A man who weighs two-hundred pounds' weight can be reported by appeal to a relation to the number two-hundred. But his weight is not a relation to the number two-hundred. The relevance of a relation to any particular abstract entity is contingent on our choice of a system of measurement. No relation between a person and an abstract entity is "the weight relation." So too, argues Matthews, there is no relation between an agent an abstract entity that is a mental state. Appeal to the abstracta is more akin to measurement (1994, 2007).

Another example: views Dennett's (1987, 2002) and Schwitzgebel's (2002, 2013) take mental states' natures to be very different than we take them to be. In Schwtizgebel's view believing p is a matter of having a bundle of dispositions that matches our cognitive stereotypes about what dispositions a person who believes p will have. Dennett's view is similar, but takes believing p to be a matter of it being profitable to attribute belief, when

adopting the intentional stance, in order to explain and predict behavior. Such views deny that the belief relation tells us about cognitive entities, like beliefs, by denying the existence of such entities. What exists and makes true belief reports is something unlike a mental state, like a pattern, property, or bundle of dispositions.

My arguments here however, contra Matthews, Dennett, and Schwitzgebel, have started from a presumption of realism about beliefs, the mental kinds that figure into thought and action, and have entered a plea of no-contest to the existence of a belief relation (for the purposes of this paper). I have argued that on that presumption of realism about the existence of beliefs and the belief relation that the belief relation is unlikely to have much to do with the nature of beliefs.

My case here does not require or entail any anti-realism about belief, beliefs, believing, believers, or the belief relation. My case also does not require any particular account of what the belief relation is, beyond just the widely accepted thesis that one of its relata is an agent. Any account of the relation's second relata or of the relation as having more than two relata are compatible with my arguments.

My case has been that through both hand changing and expiration of belief relations, there are cognitive entities or states that persist that seem to be very much like beliefs. If a belief relation were essential to the nature of belief, then these entities or states would differ from beliefs only by the fact that they do essentially involve the belief relation. What seems to make the most sense to me is to just say that they are beliefs. Denying that they are, leaves us to say they are something like *schmeliefs.* Those committed to the existence of beliefs and the essentiality of an agent relation seem to have to posit the existence of these schmeliefs, in order to make sense of the cases in the previous section. The cost of granting

that schmeliefs exist and are the things that persist in these cases, rather than beliefs, is that beliefs become eliminable from the study of the mind, in favor schmeliefs. Those who instead identify beliefs with schmeliefs both have a simpler theory and no eliminability problem. What is more, they seem to lose no explanatory power. It seems then that they are in a better position dialectically. Trying to identify mental states with relations or count the belief relation as essential to the nature of belief results in a worse theory of belief. Rejecting the importance of the belief relation to the study of belief, on the other hand, seems to cost us nothing and gives us a better theory.

I conclude then that we fare best when we deny that the belief relation has much to do with beliefs. Embracing that point puts us in a position somewhat adjacent to Matthews, counting the properties of belief reports, which have generated the relational analysis of belief, to be cottoning on to features that have more to do with how we communicate belief and engage in mindreading tasks, than they do with properties of the relevant mental phenomena. Curry seems to have something similar in mind when he urges us "to refrain from conflating attitudes of belief with intrinsic cognitive states of belief" (Curry 2021) (though I again part ways with him on the actual metaphysics).

A useful analogy for seeing how to think about the relationship between beliefs, the cognitive states of entities, and belief relations would be to think of maps. If I have a map of Utah and one of Arizona, we can distinguish the two maps by appeal to content—by what they are maps of. But if I have a map of Utah and you have a map of Utah, we can no longer distinguish them by appeal to content. But we can distinguish them by appeal to the fact that one of them is mine and the other is yours. The fact that individuation by appeal to agent and content has proved useful here by no means licenses us to infer that maps are or essentially

involve relations to agents. We could have just as easily individuated maps by appeal to the fact that one is over here and the other is over there. The fact that we can pick out an entity by a relation it bears and that doing so is useful hardly indicates anything about the modal profile of the entity or of the importance of the relation.

I believe we are in the same boat with belief reports and the belief relation. Contra Dennett and Schwitzgebel, I take our belief reports to successfully identify mental entities or states, which are in fact beliefs, more often than not. And I take the so-called "belief relation" to be a useful way of picking out such entities, precisely because it captures criteria that typically covary with the properties of beliefs that are useful to communication and mindreading practices. But I take the usefulness here to be like the usefulness of identifying a map by appeal to its relation to content and agent. I do not think there is much sense to be made of the idea that belief is a relation or that the belief relation is in any way essential to the nature of belief. Beliefs may act as realizers of belief relations, but nothing in the nature of beliefs requires any particular relation between an agent and object of belief. Appeals to relations between agents and contents to individuate beliefs are like appeals to agents and contents to individuate maps—useful for communication, but not for understanding the nature of the relevant entities.

If I am right on these points, it means that a great deal of work on the problem of meaning in mind, which takes the study of belief relations to be of great import for the study of the mind, has been a waste of time—at least, when understood as relating to the stated goal of understanding the mind. The study of belief reports is about as revelatory of the metaphysics of the mind as sandwich orders are to the metaphysics of digestion. We can

glean a little, but no serious study of digestion should take this approach. No serious study of belief should be looking at belief reports.

III. Belief and Action

It is counted a truism among most philosophers of belief that beliefs guide actions. Most of them take some kind of action guidance property to be either partly or wholly constitutive of belief.[49] While the claim is widely accepted and appeals to belief's action guiding properties figure into a number of debates in philosophy and cognitive science, the precise nature of this property and what does and does not follow from it being constitutive of belief that beliefs guide action has received surprisingly little attention in the philosophical literature. Most take themselves to have said all that needs to be said by simply echoing Ramsey's words that belief is the map by which we steer (Ramsey 1931).

Despite this lack of attention to such details, claims that certain mental states are not beliefs due to them not being action guiding at all or in the right way or to the right degree, have figured into a number of recent debates about the nature of belief and related mental states. Claims that delusions do not have the right action guidance property have been wielded to claim that the deluded to do not believe their delusions (Currie and Jureidini 2001). Claims that religious attitudes are not action guiding in the right way have been wielded to claim that many religious people do not believe what they claim to believe (Rey 2007, Van Leeuwen 2014 and Luhrmann 2018). Claims that many racial egalitarians' commitments to racial egalitarianism are not action guiding enough have been utilized to deny that they really have the egalitarian beliefs they think they have (Schwitzgebel 2010).

The aim of this chapter is to get into the weeds with belief's action guidance property and the motivations for denying agents believe what they claim to in the kinds of cases described in the previous paragraph. The picture that emerges will be one on which claims

[49] For a dissenting voice, see Strawson 1994/2009.

that a given mental state is not action guiding are much harder to defend than is usually realized. Conflicts between belief and action are common, or so I will argue, because bringing information to bear on any given situation is complicated enough that it is unrealistic to think people's beliefs can be read off of their actions with any real precision. This is not to say that beliefs do not guide action. Rather, it is to say that any plausible account of belief's action guidance property is going to need to allow that people regularly do not act in a manner consistent with their beliefs.

In section 1, I will explore the motivations for denying belief to agents in cases that have been of interest in literatures on delusions, religious beliefs, conspiracy beliefs, and racial beliefs. In section 2, I will outline two possible ways of thinking of belief's action guidance property. They both seem to be able to vindicate denials of belief in the cases mentioned above. Section 3 will then provide a budget of reasons for thinking these two accounts set an unrealistic standard for belief's action guidance property—one that would effectively entail we have few or perhaps even no beliefs at all. Section 4 will then offer some ways of solving these problems to deliver a more plausible account of belief's action guidance property. The result will be, however, that much of the evidence that philosophers have been using to deny belief in the cases described above will actually not turn out to be very good evidence at all. Section 4 will conclude the main argument of the paper. Section 5 will then dive into the metaphysics of implicit cognition. I will not take any stances on the metaphysics of implicit cognition. Instead, the section will function as a testing ground for the adequacy of the various proposals about action guidance explored in the paper.

1 – Acting contrary to belief

In recent years, cases of people acting contrary to professed belief have been of interest to philosophers and cognitive scientists. The cases that are of interest are not just any instances of people claiming to believe one thing, but not acting like it. Cases of people lying about what they believe, for example, do not seem to generate any interesting puzzles for the study of belief. The cases that have been of interest are cases where agents show many indicators of genuine belief, such as sincerity in their profession of the belief, willingness to sacrifice for the "belief" in question, betting behavior, etc.; but wherein the agent sometimes acts as if they knew the "believed" proposition to be false. Such cases have come up in the study of religion (Barrett 1999, Barrett 20014, Slone 2004, and Van Leeuwen 2014), racist attitudes (Gendler 2008, Scwhitzgebel 2010, Levy 2017), conspiracy beliefs (Mercier 2020), and delusions (Currie and Jureidini 2001 and Bortolotti 2010). Some philosophers have gone so far as to claim to have discovered novel mental states,[50] not captured in our folk psychology, by attending to such cases.

In the study of delusions, cases of *double bookkeeping* have led some to deny that delusions essentially involve a belief. Double bookkeeping cases are cases wherein the agent claims to believe something surprising, like perhaps that they are Jesus Christ, but will often act as if they knew the belief was false (e.g. they will not try to get to the other side of a body of water by walking on water or try to procure alcohol by turning water into wine or make use of their Godly powers in any way).[51] It has seemed to many psychologists and

[50] For example: Aliefs (Gendler 2008) unconscious beliefs (Mandelbaum 2013, 2016), patchy endorsements (Levy 2015), and religious credences (Van Leeuwen 2014). There are also ongoing debates over whether delusions are beliefs (Bortolotti 2010, 2011) or a distinct mental kind (Currie and Jureidini 2001), which centers around the phenomenon of double bookkeeping, wherein delusional patients act contrary to their professed delusions. All told, there is no shortage of novel mental kinds being posited as a result of these kinds of cases.

[51] See Rokeach 1964 for a case study involving the behaviors of people who claimed to believe they are Jesus Christ. Rokeach also comes down on the side of doxasticism about delusions within the book.

philosophers as if the person suffering from a delusion has two maps (or sets of books) they are steering by—one featuring the delusion and other attitudes connected to it, and the other an ordinary map without the delusion or any other outlandish beliefs.

Similar phenomena have cropped up in the cognitive science of religion, where cases studied under the heading of *theological incorrectness* involve agents with a strong commitment to their religion often acting or reasoning as if they knew the religious propositions they claimed to believe to be false. A Christian might say they believe God is omniscient, but act as if God needs to be informed of recent developments in their lives; or she might believe God is omnipotent, but act and reason as if God can only do one task at a time (Barrett 1999 and Slone 2004). Neil Van Leeuwen has recently argued that such cases show us that there are really two different kinds of mental states that we use the word "belief" to pick out, and that religious agents often stand in a relation to religious propositions that is quite different than what we would normally expect from a belief. The religious attitude is a different mental kind entirely—one he calls a *religious credence* (Van Leeuwen 2014).

When dealing with racial beliefs, cases of people priding themselves on their racially egalitarian beliefs, being willing to join the picket line, being willing to argue articulately and passionately for their beliefs, but nonetheless being surprised when a black person says something smart or frightened when a black person is walking toward them at night, have been used to deny that the agents in question really hold racially egalitarian beliefs.[52]

Such cases are interesting, because they seem to generate inconsistent triads roughly fitting the following schema:

[52] See Schwitzgebel 2010 for discussion

(IC1) Beliefs are action guiding

(IC2) The mental states satisfying description X are not action guiding

(IC3) The mental states satisfying description X are beliefs

When we fill in the schema with a case from one of these literatures, we get an inconsistent triad. The "mental states satisfying description X" can be the religious attitudes studied in the theological incorrectness literature, delusions, certain racial attitudes, conspiracy beliefs, or any other mental state we would be tempted to call a belief, but which we might also be tempted to say fails to guide actions at all, enough, or in the right way.

When the schema is filled in, the resulting propositions can be used to construct arguments with any two of the propositions being used to form a valid argument that denies the third. Those denying belief in the cases mentioned above take themselves to have strong enough reasons for holding onto IC2 and IC1, and thus deny IC3, saying the states in question are not beliefs at all.

2 – Practical Setting Independence and Action Consistency

From here on out, I will call the kinds of cases we are dealing with *belief action conflicts*, or *BACs* for short. When we have a BAC, we have a puzzle to be solved in the form of the inconsistent triad fitting the schema provided in the previous section. We either need to deny that the state in question is a belief or make sense of belief's action guidance property in such a way as to understand how the state in question could be a belief, despite the BAC.[53]

[53] The other alternative is to deny that beliefs guide actions. This paper will not explore that possibility.

In order to understand when a BAC should lead us to deny that the agent believes, we need to get a better grip of what it is for beliefs to guide action. I'll take as a starting point, the property of *practical setting independence*, an account of which was offered by Neil Van Leeuwen as a way of understanding belief's action guidance property. Here is Van Leeuwen's account of practical setting independence:

> *Practical Setting Independence:* S's attitude X is practical setting independent iff X serves as a premise in practical reasoning in all practical settings where its content is relevant to S's actions (Van Leeuwen 2014).[54]

Practical setting independence is supposed to be a step forward in thinking about action guidance, because it specifies conditions under which beliefs guide actions. My belief that Tony Danza starred on *Who's The Boss?* does not need to guide my actions when I am driving to the beach in order to be a belief, since its content is not relevant. If on the other hand, I am trying to collect the autographs of everyone who was on the show, that belief should figure into my practical reasoning.

I have a serious worry about whether serving as a premise in practical reasoning could possibly be a constitutive feature of belief. I call this worry the *Too Many Beliefs Objection*. It seems straightforwardly true that some beliefs do not serve as premises in practical reasoning when their content is relevant because other beliefs with similar content play that role instead. Say for example, that I am at a conference with Daniel Dennett, and I want to shake his hand. I believe he is between Aaron Zimmerman and Michael Gazzaniga.

[54] Note: this is my wording. Van Leeuwen actually defines practical setting independence in the following terms: "(1) A cognitive attitude x is practical setting independent if and only if x guides behavior in all practical settings in which x's content is relevant to the agent's behaviors. (2) A class of cognitive attitudes X is practical setting independent if and only if X is employed in guiding action in all practical settings." He elsewhere in the same paper explains that he is understanding guiding action as serving as a premise in practical reasoning. The definition I attribute to him seems more in line with the view he actually argues from than his own definition.

I also believe he is to the right of Aaron Zimmerman. I also believe he is to the left of Michael Gazzaniga. I also have a mental map of the room and I have beliefs about his location on that map. On any account of how to individuate beliefs, these beliefs are likely going to come out as distinct beliefs. And yet any one of these beliefs could serve as a premise in practical reasoning to the exclusion of any of the others, despite the fact that all of their contents are relevant to the location of Daniel Dennett. Citing any one of the beliefs could rationalize my behavior in navigating the room. But there is no reason why each needs to have served as a premise in practical reasoning when their content was relevant.

There is nothing particularly special about this case. In most circumstances there are a number of beliefs that we could cite to explain a person's behavior equally well. We are *normally* in the circumstance of having more beliefs than we can use as premises in practical reasoning. So, it seems pretty routine for a belief's content to be relevant and for it to not serve as a premise in practical reasoning, because the agent has other beliefs that can do the job equally well.

It may be that there is a way of futzing with the wording of *practical setting independence* to get around the too many beliefs objection. But exploring the matter would derail the project of the paper too much. Instead, I will offer a second possible account of action guidance. Throughout the paper, I will make mention of both accounts, in case the Too Many Beliefs Objection can be answered.

> *Action Consistency*: S's attitude X is action guiding iff S's actions are consistent X in the actual world and in all nearby worlds

In order to understand this new account, it will be useful to focus on two elements of it: consistency and possible worlds. The consistency in question is difficult to adequately

characterize, but easy enough to recognize. A person who believes his keys are in the living room drawer is acting consistently with his belief when he goes to the drawer to retrieve the keys. A person who proclaims to believe in racial equality, but actively supports the KKK is acting inconsistently with a belief in racial equality. The inconsistency here is not a contradiction in formal logic, of the form p and ~p. But it is *some* kind of inconsistency.

The possible worlds language is included in order to rule out propositions that are consistent with an agent's behavior in the actual world, but which are not beliefs. Take for example an infinite series of propositions of the form "x is the number of hills on Pluto," where for every number greater than zero, there is a proposition in the series including that number as a value for x. I do not have any beliefs about the number of hills on Pluto, but it seems as though all of the actions I take in my actual life are consistent with a great number of these propositions. So, in order for us to not have to say that I stand in an action guidance relation with all of these propositions, we will have to appeal to some counterfactuals—to what I would do under different conditions. If I *were* asked how many hills there are on Pluto, I *would* answer that I do not know. Were I to be placed in the position of having to gamble on whether the number was even or odd, I would have no grounds for betting on one or the other. So, I would not count as believing any of the propositions on that list, because were they to be relevant I would not act in a manner that is consistent with a belief in any of them.

Action Consistency gets around the too many beliefs objection with ease, because no matter what belief I use to navigate the room to get to Dennett, my actions will be consistent with all of my beliefs.

Action Consistency may however strike many as being too passive to be an account of action *guidance.* A critic may wish to say that what we want is an explanation of how beliefs interact with the cognitive system in order to actually produce actions. This is something Practical Setting Independence offers, but Action Consistency lacks.

My own take on this worry is that Action Consistency allows that the action guidance property may be realized in multiple ways. Sometimes it may be a matter of beliefs serving as a premise in practical reasoning. Sometimes what is relevant is that beliefs may guide action by taking certain actions "off the table." In other words, if we return to the hand shaking example, the route I take to Dennett is explained in part by it being one of the routes that are available. And the other beliefs involved, despite not being premises in practical reasoning, do in fact narrow down the space of possible actions I consider.

This may lead us to want to consider a third possible account of action guidance, that looks something like this:

> *Possibility Narrowing:* attitude x guides S's actions iff x narrows the space of possible actions.

While I acknowledge there is something attractive about Possibility Narrowing, it shares similar problems with Practical Setting Independence. Too many beliefs do not narrow the space of possible actions. Some beliefs expand the menu of possible actions when formed, such as when you become aware of an ability you did not realize you had. Some beliefs do not change the options available to you at all.

While I acknowledge that there are further questions about the adequacy of Action Consistency, Possibility Narrowing, and Practical Setting Independence, they are questions that will need to be pursued elsewhere, as they are orthogonal to the overall project of this

paper. For now, I will proceed with both Practical Setting Independence and Action Consistency in hand as possible accounts of action guidance. The next section will aim to show that neither one is a realistic standard for action guidance, and for pretty much the same reasons. Section 4, will offer fixes for both.

3 – Acting Contrary to Belief

The accounts of action guidance offered in the previous section seem to support denial of belief to agents in BACs. They represent the picture of action guidance that seems to be presumed in the argumentation denying belief when dealing with BACs. This section will show, however, that if either one is right, we would have little to no beliefs at all.

I will proceed in the subsections below by moving through a series of problems and problematic cases for the accounts, which suggest that the accounts need to be much weaker to be plausible. In section 4, I will turn to the issue of providing fixes. Here I will provide only problems.

3.1 – Non-retrieval cases

The first kind of example I will wield against the action guidance accounts offered in the previous section is what I will call a *non-retrieval case.* These are cases where information is not activated prior to acting in order to guide action. Consider a person who has decided to quit caffeine. The next morning, she gets up, goes for a jog, stops at a convenience store to catch her breath, habitually buys an energy drink, and drinks half of it before recalling that she is not supposed to be drinking caffeine. In this case, she had a number of beliefs whose content was relevant that neither served as premises in practical

reasoning nor were consistent with her actions. She believed caffeinated sodas were no longer appropriate, that she should not drink caffeine, that the caffeinated drink she had the day before was her last one for quite some time, etc. But none of these beliefs showed up in time to guide action. They are beliefs all the same. So, it cannot be a constitutive feature of beliefs that they serve as premises in practical reasoning whenever their content is relevant or that actions in the actual or nearby worlds need to always be consistent with them.

A second type of example of non-retrieval cases would be cases where decisions have to be made quickly. Sometimes, we find ourselves in the position of saying, "if I would have had more time to think about it, I would have done y instead of x." What else is this but a profession that there was content relevant to guiding behavior that did not guide behavior? If that is right, then it cannot be the case that beliefs guide action whenever their content is relevant or that they need to be consistent with beliefs.

I mention both examples of non-retrieval cases, because it is plausible to think of the first example as a malfunction. It is not plausible to think of the second example as a malfunction. It will be useful to keep that fact in mind, when we are considering possible fixes. Everything can be functioning optimally, without the mind being able to solve a problem in a small enough window of time. To say otherwise would be like saying an inability to run a 3 second mile is a malfunction of the musculature system. Correct functioning does not amount to omnipotence. So, however we address these problems, focusing on malfunction is unlikely to be a solution.

3.1 – Non-inference cases

The second type of problem case I'll wield are what I call *non-inference cases*. These are cases where acting requires drawing some inferences and the agent fails to do so. Acting upon a belief often requires drawing some inferences about how or whether a piece of content bears on the situation in question. These cases are especially relevant to the cognitive science of religion, where one explainer of BACs is that the relevant religious beliefs have *counterintuitive content.*[55] The basic picture on offer from these cognitive scientists is that we have something like an intuitive metaphysics—a set of commitments about how the world works and how things in the world are typed. We are good at drawing inferences and reasoning quickly and easily about things that match our intuitive metaphysical theory, but when they do not match our intuitive metaphysics—when things are *counterintuitive*—we are prone to fail to draw inferences. Thus, a belief's content may be relevant, but may not serve as a premise in practical reasoning, because the agent has failed to draw the necessary inferences about its relevance to the behavioral context.

Other causes of non-inference may be a lack of interest in the subject matter, being under cognitive load, or having low levels of cognitive ease with the subject matter. Whatever the root cause, failing to infer can cause failures to act in accord with beliefs or to take a belief as a premise when its content is relevant.

3.3 – Non-ability cases

The next type of problematic cases are what I'll call *non-ability cases*. These are cases where there is a failure to act consistently with beliefs due to a lack of ability. The

[55] See Barrett 2004 for an overview.

most common kind of case in this vein is a case of akrasia. The agent believes they should do something, but lacks the strength of will to do so.

There are, however, a wide range of cases that do not seem to be problems of a weak will. A person with ADHD may fail to act on their beliefs, because they are having a difficult time sustaining attention on the task or because they are hyperfocused on some bit of information to the point where they have failed to incorporate other information that they have available. This person is not having a will problem, they are having an ADHD problem. Similar issues occur when thinking about people dealing with illnesses. Further, there are also some tasks that require enormous feats of will. A person may have insufficient will, but a lack of superhuman will power is certainly not a *weakness* of will. So, non-ability problems include akrasia cases, but also include a wide range of problems not related to the will.

Such cases may result in failures to take a belief as a premise in practical reasoning, such as the hyperfocused ADHD person who is having trouble attending to some of her relevant beliefs. Or they may result in a person acting inconsistent with their beliefs, as in the case of the akratic person.

3.4 – Mistakes of practical reasoning

A kind of case that may present a challenge to Action Consistency, but not Practical Setting Independence would be *mistakes of practical reasoning.* If an agent takes a belief as a premise in practical reasoning, but makes some mistake in practical reasoning in the formation of an intention or the production of action, the agent may not act in accord with their belief. There is some controversy in the philosophy of practical reasoning about

whether mistakes of practical reasoning are even possible[56] and there is no real agreement about just what practical reasoning is. So, it is difficult to discuss the matter effectively. But if mistakes of practical reasoning are possible, they plausibly would be cases wherein a belief serves as a premise in practical reasoning, but wherein actions are not consistent with belief.

3.5 – Ignorance of Relevance

A final concern I will raise about our theories of action guidance is *ignorance of relevance.* The most straightforward way of generating such a case is to consider the accounts' relations to problems of content. There is no agreement about what the correct theory of content is, but it seems as though broad content accounts will spell trouble for both of our accounts. On a Millian theory of content, for example, the sentences "Mark Twain should be arrested" and "Samuel Clemens should be arrested" both have the same content. Suppose a police officer pulls over a Mark Twain-ish looking person, asks for ID, reads "Samuel Clemens" and lets the fellow go, despite having a belief that he would report as "Mark Twain should be arrested." The content of that belief was relevant, but it failed to serve as a premise in practical reasoning when he was deciding whether or not to arrest the man and his actions were inconsistent with his belief's content. Similar cases can be constructed for descriptivist theories of content using "Mark Twain" and "The author of Huckleberry Finn."

[56] See Lavin 2004 for exploration of this issue

Advocates of narrow content may not be too worried about such concerns. But it would be nice if whatever fixes we propose could sidestep such concerns about content entirely, so as to not wed our theory of belief to the embattled notion of narrow content.

Cases of ignorance of relevance can plausibly occur in other ways as well, for example, by failing to recognize a person or object for what it is (e.g. you fail to recognize a bomb, thus failing to bring your bomb related beliefs to bear on the situation), or because recognizing the relevance requires a deeper understanding of the material than the agent possesses. However, such cases are generated, if it is possible to be ignorant of the relevance of a piece of believed content, the belief will not serve as a premise in practical reasoning when it is relevant and actions taken may not be consistent with the belief.

3.6 – The takeaway

I've offered here a number of problems for our accounts of action guidance from section 2. I do not think that any of them are particularly challenging for the notion that beliefs guide actions. However, they are all problems for the accounts of action guidance that are plausible candidates motivating denials of belief in BACs from the literatures on delusions, religious beliefs, conspiracy beliefs, and racial beliefs.

4 – Strategies for dealing with section 3 problems

We have seen that the accounts from section 2 run into some damning problems. Neither of the accounts seems to be a plausible account of belief's action guidance property. This section will consider ways of handling such cases.

4.1 will address the problem by adding condition clauses to the accounts in order to narrow in on the conditions under which we can either expect beliefs to serve as premises in practical reasoning or expect consistency between belief and action. 4.2 will consider omitting the condition clauses, in favor of a teleofunctionalist account of mental states. 4.3 will consider ignoring the problem entirely by appealing to implicit ceteris paribus clauses. I won't take any stance here about how one *should* address the problems. Rather, I will show in 4.4 that no matter which of these options we go for, we will be stuck with the conclusion that actions running contrary to belief are commonplace and BACs are often not very good evidence for denying belief.

4.1 – Adding condition clauses

Practical Setting Independence pushes forward the discussion on belief's action guiding properties in two ways. One way is that it offers an account of what action guidance consists in (serving as a premise in practical reasoning). The other way is that it specifies the conditions under which this occurs (when the content of the belief is relevant). As we have seen, the account fails spectacularly, because it is simply far too unrealistic of a standard to expect beliefs to guide action whenever their content is relevant. Bringing information to bear on a situation when its content is relevant is complicated task, requiring a suite of competences, capacities, and conditions that are not universally realized in all action contexts.

But Practical Setting Independence itself does suggest a way forward. We can just add more condition clauses to specify the conditions under which action guidance can be

expected to occur. We can apply the strategy for our two accounts from section two, resulting in the following two accounts:

> *Conditioned Practical Setting Independence:* S's attitude X is practical setting independent iff X serves as a premise in S's practical reasoning whenever (a) X's content is relevant to S's behavior, (b) S recognizes the relevance of X to her behavior, (c) X is successfully retrieved in time for S to bring X to bear on her behavior, and (d) acting upon X is within S's capabilities.
>
> *Conditioned Action Consistency:* S's attitude X is action guiding iff S's actions are consistent with X, in the actual world and in all nearby worlds, whenever the following conditions are met: (a) X's content is relevant to S's behavior, (b) S recognizes the relevance of X to her behavior, (c) X is successfully retrieved in time for S to bring X to bear on her behavior, (d) acting upon X is within S's capabilities, and (e) S has made no error in practical reasoning involving X.

Our new accounts seem much better at dealing with the cases from section 3. The a-clauses in each are Van Leeuwen's content condition. It is not obvious that Conditioned Action Consistency needs an a-clause, but it seems to do no harm leaving it in.

Our b-clauses deal with ignorance of relevance cases and non-inference cases. C-clauses deal with deal with non-retrieval cases. D-clauses deal with non-ability cases.

Conditioned Action Consistency has an e-clause that is not shared with Conditioned Practical Setting Independence. The e-clause is there to deal with mistakes in practical reasoning, which generated problems for Action Consistency, but not Practical Setting Independence.

There may be room for some debate about whether these are the optimal condition clauses, or whether there are further cases I have not considered in this paper that would warrant further condition clauses. As far as I am concerned, all of that is welcome and not a threat to my overall project here. My project here is to narrow in enough on what an adequate account of belief's action guidance property is going to have to look like to show

why we should think that BACs are common and often not very good indicators of belief. Questions of whether I have the optimal condition clauses are orthogonal to that project, so long as I have provided as I have made enough progress to make it clear why that is the case. (More on why that is supposed to be the cases in 4.4.)

4.2 – Teleofunctionalism

A second route for dealing with the problems from section 3 is to adopt a teleofunctionalist metaphysics of the mind. Doing so would obviate the need for any condition clauses and could allow our accounts from section 2 to stand as written, without correction.

On a teleofunctionalist approach, what is constitutive of a mental state is not its actual function but rather its normal function or ideal function.[57] Going with normal function would not be of much help, since non-retrieval cases do happen as part of normal function, when actions have to be performed under severe time constraints.[58] We sometimes have to act more quickly than we can bring information to mind, resulting in failing to bring information to bear on a situation.

Going for ideal function however would be of help. Ideally, information is stored so that it can be used when it is relevant. So, a teleofunctionalist reading of Practical Setting Independence and Action Consistency render them both plausible.

4.3 – Implicit ceteris paribus

[57] See Godfrey-Smith 1998 for an exploration of teleofunctionalism and the metaphysics of the mind. See Miyazono 2019 for an exploration of this approach to belief and delusion.

[58] This issue was explored in 2.2 of this chapter.

Many follow Fodor in taking it to be the case that all generalizations in philosophical psychology have an implicit ceteris paribus clause (Fodor 1987). Someone who thinks this might criticize my work here, saying Practical Setting Independence already had an implicit ceteris paribus clause, and all I've done is unpack the conditions for the ceteris paribus clause. But I have not exposed any problems for it as an account of belief's action guidance property.

My first reply to this is that even that is the case, it is still important to unpack the ceteris paribus clause if we are to understand the conditions under which beliefs guide actions and assess the evidential value of actions that run contrary to professed beliefs when attributing and denying beliefs or when doing the metaphysics of the mind.

My second reply to it is that, as we'll see in 4.4, once we have spelled out these additional conditions, the evidential value of BACs is significantly undermined. And that is reason enough to think that people arguing from BACs to a denial of belief are not presuming these conditions to implicitly be included already.

4.4 – The evidential value of BACs

I've explored here three ways of handling the problems I raised in section 3. One approach involves adding condition clauses. Another approach reconceived the project of doing the metaphysics of the mind as involving giving functional specifications of mental states' functioning in ideal minds (teleofunctionalism). The last approach said we should think of generalizations in the philosophy of mind as including implicit ceteris paribus

clauses, and think of the condition clauses added in 4.1 as having implicitly been there all along.

Whatever approach one takes, the fact remains that the problems raised in section three show that conflicts between beliefs and actions are commonplace. It also means that for BACs to have evidential value in denying beliefs to agents, we have to have good reason to think that the condition clauses are satisfied. Bringing information to bear on any given situation requires a suite of competences and conditions that are often not satisfied.

In the study of delusions, for example, cases of double bookkeeping involve agents that are pretty far from being ideal reasoners. There seems to be no a priori reason to presume that they have successfully considered the full implications of their beliefs, or capably attended to how to implement a behavioral repertoire that reflects their attitudes. When dealing with a delusional belief, such as that one is Jesus Christ or that the leg attached to one's own body is not one's own, it is difficult to know what just what the implications are or what a good course of action would look like. We should expect low levels of cognitive ease when dealing with such circumstances and for the intuitive metaphysics built into our cognitive architecture to be nonplussed. Similar points can be made for conspiracy beliefs, religious beliefs, and racial beliefs.

Acting on one's beliefs is not so easy as it seems. The fact that we do it so easily and effortlessly when dealing with situations where we have a high degree of cognitive ease, a high degree of compatibility with our intuitive metaphysics, and the examples of countless others with similar beliefs to work with should not lead us to think that when dealing with attitudes whose content lacks one or more of those properties that it will be easy enough to expect consistency between action and belief.

5 – Implicit cognition

Section 4 completes the main argument of this chapter. This section turns to the metaphysics of implicit cognition. The reasons for doing so are twofold. First, BACs occur in the literature on implicit cognition, where they have been used to argue in favor of the existence of implicit attitudes. Second, the study of implicit cognition provides more grist for the mill for showing the inadequacies of Practical Setting Independence and Action Consistency, while also demonstrating the adequacy of their Conditioned counterparts. In other words, this section can be considered to be something like a field test for the conclusions we have reached so far.

Many BACs are cases wherein implicit attitudes run contrary to reflectively endorsed attitudes. Not all such cases pose a threat to our accounts. For example, one of Gendler's examples is of a person on a glass bridge, quaking when they look down. The person got on the bridge because they took it to be safe, but the quaking seems to suggest that they do not really think it is safe (Gendler 2008). But, in such a case, the quaking is a reflex and thus not an action. It thus does not pose any threat to our accounts of action guidance, because the agent is not performing any action contrary to their professed belief that the bridge is safe.

But there are still troubling cases wherein implicit cognition seems to guide action in a way that is contrary to the agent's professed beliefs. We might imagine a liberal professor, who champions racial equality in her lectures and joins the picket line in the fight for racial equality, but nonetheless is always surprised when a black person says something smart.[59] Surprise is again not an action, but a reflex. It is nonetheless typically thought to be an

[59] This example is very close to one from Schwitzgebel 2010

important indicator of agents' attitudes. Further, from the fact that she is surprised, we can expect her behavior to be affected in other ways. When she grades a black student's paper, she is less likely to be disposed to be charitable about mistakes. When she is looking at students' hands, deciding who to call on, if she is looking for a smart person to get discussion going, she will likely be less apt to call on black students.

Such a case seems best explained by an implicit attitude. There are competing accounts about what implicit attitudes are. Traditionally, within psychology, implicit attitudes have been seen as associations. If that is what is going on here, then what might explain the professor's actions is that seeing a black person may prime associations with behavioral routines and dispositions to judge in certain ways that are deemed appropriate for dealing with unintelligent people. While priming these associated routines does not determine that the agent will employ such routines, the likelihood of her treating black people as less intelligent will be significantly increased, because those are the behavioral routines that are readied.

Many philosophers have been skeptical that a mere association could do the explanatory work here. Gendler has argued that implicit attitudes are *aliefs,* a hodgepodge kind of mental state constructed out of associations, representations, and affective states. The agent thus alieves black people are less intelligent but believes that they are just as intelligent as any other race (Gendler 2008). Mandelbaum has argued that the implicit attitudes are just beliefs, albeit unconscious ones (Mandelbaum 2013 and 2016). On this account, the agent has an unconscious belief that black people are less intelligent but a conscious one that they are just as intelligent. Levy has argued that they are a different kind of mental state altogether, which he calls *patchy endorsements* (Levy 2015). According to

Levy, these states are like beliefs, but typically unconscious and less representationally rich than beliefs. The agent thus has a patchy endorsement of the racist attitude and a belief in racial equality.

Whatever the correct account of implicit attitudes may be, they all spell trouble for our accounts from section 2. The professor's racial egalitarian commitments are relevant to how the student is treated, but it is not serving as a premise in practical reasoning and her actions are not consistent with her beliefs. If we instead say she believes that black people are less intelligent, then that belief is not guiding action when she is out on the picket line or giving passionate lectures about racial equality. No matter what belief we attribute to her, we have a *prima facie* counterexample to our accounts from section 2.

If Mandelbaum is right, on the other hand, then agents have contradictory beliefs, since the unconscious ones contradict the conscious ones. Thus, differing contradictory beliefs are guiding different actions at different times. Beliefs are thus not guiding action whenever their content is relevant, because sometimes a belief is relevant but a belief to the contrary is instead guiding action.

If the psychologists are right, then negative associations can sometimes prompt behavioral routines without the relevant beliefs being involved at all. A network of associated information or behavioral routines is used to handle the situation, which can omit relevant beliefs entirely. Thus, even when a belief's content is relevant, it will fail to guide action. (This would technically make it a non-retrieval case.)

If Gendler (aliefs) or Levy (patchy endorsements) are right, then it is a mental state of a different kind guiding action, serving as a premise in unconscious practical reasoning or causing action in some other way. However we look at it, beliefs are not guiding action

when their content is relevant and actions are being performed that are inconsistent with agents' beliefs.

There are differing accounts about what we should say about the professor's putative egalitarian belief as well. And how we move forward on belief's action guiding properties will depend on what we have to say about this as well.

On some accounts, belief is so behaviorally demanding that we should say that the professor does not believe in racial equality after all (e.g. Schwitzgebel 2021). These little moments of surprise are more telling than all of the time she spent on picket line or trying to persuade people of the truth of racial equality. Schwitzgebel (2010) calls such views *anti-judgment views.*

Other accounts say we should explain what is going on here as a case of shifting belief. Sometimes the agent believes in racial equality and sometimes she does not. Call this *the shifting view* (Rowbottom 2007).

Other accounts say that we should follow the agent's conscious acts of judgments, in order to settle what she believes. If we were to point out to the professor that she was treating black people differently, she would likely change her behavior to accord with her commitments to racial equality, since that is what she judged to be accurate on reflection. It seems unlikely to go in the other direction—of having her give up on arguing for racial equality, because she is sometimes disposed to unequal treatment. Not all actions accord with beliefs, but when agents recognize that their actions do not accord with their beliefs, they change their actions to bring them in line with their beliefs. Call such views, *pro-judgment views* (Zimmerman 2007 and Gendler 2008). Such accounts find an easy fit with

Gendler and Levy's views, as well as the psychologists views of the attitudes as being merely associations.

I count the anti-judgment view as a non-starter. If the anti-judgment views were right, then we would have almost no beliefs at all. As detailed in sections 3, it is just plain unrealistic to think anyone's beliefs and actions are in perfect accord. Bringing information to bear on a situation when that information is relevant is difficult work for a cognitive system and no one satisfies the ideal conditions. So I take the position to just be a non-starter.

The shifting view and the pro-judgment view fare much better. I am not convinced our natural language uses of the term *belief* or intuitions can settle the matter between the two. When lay people talk about implicit attitudes, they use terms like "you believe on some level…", suggesting that belief talk is taken to be just as appropriate for implicit attitudes that do not accord with our judgments as it is when used to talk about things we judge true.

On the other hand, if we pointed out to our professor that she is behaving in a racially prejudiced way, we would expect, if she really believed in racial equality, that she would change her behavior. And it would seem to make perfect sense to explain her behavioral change as being a product of the fact that she *really believes* in racial equality.

While I favor the pro-judgment view, it is beyond the scope of this chapter to try and adjudicate this matter. What is at issue here is what the implications are for beliefs' action guiding properties. If the shifting view is right, then unconscious states with the content ~p and conscious states with the content p shift in their status as beliefs, depending upon what the agents are acting on. Technically, this does not run afoul of either of our accounts from section 2. Take for example a conscious state with content p that acts as a belief. When the

agent acts on their unconscious state with the content ~p, the p state ceases to be a belief. When they act on the conscious p state, the ~p state ceases to be a belief. So, the states guide action whenever they are beliefs with relevant content. But the states do not guide action whenever their content is relevant. This may seem a bit contradictory, but it is worth noting that the same representation can be a belief, a doubt, a hope, etc. So, a representation's shifting status from being a belief to some non-belief state cannot count against either of our accounts. So, if the shifting view is correct, then we have no challenge to any version of the action guidance property we have considered here, though we once again arrive at an understanding of action guidance that undermines any case that BACs are good grounds for denying belief.

If the pro-judgment view is correct, then we have challenges to Practical Setting Independence and Action Consistency. Beliefs are not guiding action when their content is relevant. Either associations are prompting behavioral routines that are not taking into account relevant beliefs, or some other unconscious representation is guiding action. Either way, we have relevant content without the beliefs serving as premises in practical reasoning and we have inconsistency between belief and action.

The Conditioned counterparts fare better, however. It seems as though what is going on in the case of the professor is that she is not recognizing how her racial equality beliefs bear on the situation.[60] Thus, the case is plausibly handled by our b-clauses and c-clauses. When teaching a class, a person has to think on their feet and make a lot of decisions quickly. These are precisely the kinds of circumstances that a person can fail to act on their

[60] She could of course also just be a closet racist. We cannot rule that out. But my point here is that nothing described in the case does much to support that conclusion. We would need further evidence to support that conclusion.

beliefs, because they are not able to recognize the relevance of a piece of content or successfully retrieve it in time for it to guide action. They are also not likely to be able to attend to how successfully the many quick successive decisions accorded with their actions. They are thus much more likely to have a failure to bring all the information they possess to bear on the situation and much less likely to recognize the failure to bring such information to bear as well.

An additional fix that we may want to consider is Zimmerman's attention and control clauses (Zimmerman 2018). Zimmerman is an advocate of the pro-judgment view and counts a state a belief if it guides relatively controlled and attentive actions. Actions that occur when we aren't paying adequate attention to what we are doing may not always be in line with what we judge true.

I don't have any objection to adding these clauses to Practical Setting Independence or Action Consistency—as I said, I am sympathetic to pro-judgment views. I do not think however that they are adequate on their own without the clauses added in the Conditioned counterparts, for reasons already covered in previous sections (e.g. you can be attending to what you are doing, but still fail to recall information or draw appropriate inferences). Adjudicating the matter fully however will not be possible here. I mention it only to make clear that these debates over how to think of belief and its relation to implicit cognition may well require further revisions to our accounts of action guidance.

Section 6 – Conclusion

I've considered here two broad ways of approaching action guidance. One of them is counterfactually robust notion of consistency between belief and action. The other is beliefs

serving as premises in practical reasoning. Whichever of these accounts we opt for, it seems that we are going to have to allow that people acting contrary to their belief is commonplace.

The one problem I haven't solved is the *too many beliefs* objection for Practical Setting Independence and its Conditioned counterpart. It may well be that all that is needed for it is some cleverness in thinking through a way to make it counterfactually robust. But I have my doubts about its repairability. And for that reason, I favor Conditioned Action Consistency. But if the problem could be solved, I would count both accounts adequate. (No doubt, the reason why actions tend to be consistent with beliefs is often as a result of them being used as premises in practical reasoning.)

Along the way, we have also gained some resources for assessing the evidential value of BACs. No doubt, some cases of people acting contrary to their professed beliefs should lead us to deny that they believe what they profess. People do lie, both to themselves and others, about what they believe. But we have also seen that actions that run contrary to beliefs should be commonplace. Bringing information to bear on any situation where the content is relevant is a complex task. When dealing with religious people, delusional people, conspiracy theorists, or anyone else failing to act in accordance with their beliefs, we should not be quick to deny they believe what they are saying. Failing to bring information to bear when it is relevant is common thing. No one lives up to their beliefs all the time. We sometimes use the maps we steer by poorly and we often have to act too quickly to even give the map a good look. We should thus instead see acting consistently with one's beliefs not as something that is constitutive of belief, but rather as an achievement—something we aspire to, but often fail to achieve.

IV. The Belief-Emotion Nexus

The aim of this essay is to theorize the intersection of belief and emotion—what I call *the belief-emotion nexus*. The story I have to tell about that nexus is one that I suspect will strike some as terribly obvious and others as quite shocking. The reason for the disparity in reactions I expect is not the usual facts about how philosophers cannot agree on anything. Rather, the reason, as I see it, is that much of the picture on offer here is operative in our tacit understanding of the belief-emotion nexus, which can be seen at work in the advice we give to each other, the way we make decisions, the stories we tell, and the media we consume. Many will hear it and find themselves in it quite readily. But philosophical debates about belief and emotion have so thoroughly neglected so many elements of this story that I suspect that many theorists will be so entrenched in the positions they have taken as to not be able to readily accommodate much of what I have to say here.

The story I have to tell about this nexus is a synthesis of recent work on affective neuroscience, which has undergone significant revolutions in recent decades; error-management-theoretic evolutionary psychology, which takes the avoidance of costly errors to be a factor in evolution (Haselton and Buss 2000, Barrett 2004, Maij et al 2019); and some hard-won lessons from the philosophical study of the mind. Elements of this story can be found in the work of neuroscientists Ralph Adolphs and David Anderson (2018), in a recent popular science book on emotion by Leonard Mlodinow (2022), and in the blossoming evolutionary psychology research paradigm of error management theory. I suspect that such thinkers may find much of my work here a natural extension of theirs, but it is distinct enough from what they have so far said that it is conceivable that they could disagree with almost all of it. Some of the elements of this story are explicitly present in

their work, others strike me as being tacit in that work, some small amount of what I have to say is at odds with some of that work, and much that I have to say goes far beyond that work.

At the core of my analysis of the interactions between belief and emotion will be an error-management-theoretic analysis of emotions. It is an analysis of emotions in only the loosest terms. It does not seek to provide a definition of emotion. It uses "emotion" loosely, including a broad range of affective states, which many may wish to not call emotions for various reasons, and is compatible with a wide range of views about the metaphysics of emotions and other affective states. So it is not a traditional philosophical analysis of emotion so much as it is a look at emotion through the lens of error management theory, with an eye to highlighting salient features that are relevant to the belief-emotion nexus.

The story I have to tell, in brief, is that affective states are constantly shaping our practices of belief formation and revision in a way that is aimed at avoiding salient but costly errors. The errors in question are not errors of truth or falsity of belief, but errors that could lead to death, injury, or missing out on important species goods, like food, mating, security, or status. Emotions pattern our attention toward goals, in an error avoidant way, raising our epistemic vigilance with respect to certain kinds of belief, while lowering our epistemic vigilance with respect to others. Emotions are or involve affective routines that allow for greater flexibility than reflex actions, but which still allow us to act quickly to avoid dying or missing out on species goods. They thus fill a niche in our cognitive repertoire to correct for the inflexibility of reflex actions on the one hand and the slowness of higher-level reasoning processes, which are not always decisive anyway. They do this by guiding action in a way that defaults to the avoidance of costly errors, where mere appeal to

the evidence might not be sufficient to guide action quickly enough or to successfully avoid costly errors. Thus, emotions systematically bias our thinking toward beliefs and behavioral routines aimed at ensuring we are consistently (though not unfailingly) steered toward value and away from unnecessary risk.

I posit a two-step error avoidance routine in neurotypical human cognition. It can be accommodated on a variety of frameworks for human cognition. But the basic idea is that since emotional thinking is error avoidant and biases us away from strictly evidential norms, it is prone to a kind of *error pileup.* Being error avoidant often is not being truth acquisitive. If too many false beliefs pileup, the organism is at risk for no longer being able to detect salient costly errors, because their understanding of the world is so mistaken. What is error avoidant in short term decision making is not always error avoidant in the long run. Deliberate, slow thinking serves as a corrective, aimed at correcting errors and writing wrongs, when cooler headed thinking is at work. The second step of error avoidance is akin to checking the work that the emotions have done.

I argue that failure to emotionally regulate and engage such deliberative processes can cause a specific kind of runaway error pileup, which I call *Emotion-Belief Cycling* (EB-cycling), wherein the effects of some particular emotion on belief formation and revision makes agents prone to beliefs of a certain type, but failure to downregulate and engage deliberative processes (to check the work) causes those beliefs to "entrench", resulting in making one more prone to experiencing that emotional response (since those beliefs rationalize that emotional response), leading one to form more beliefs of that type, leading to more experiences of the emotion, cycling on and on to some very weird places. I argue that

much of political polarization, radicalization, and believing badly can be explained by EB-cycling.

1 – Error Management Theory

Error management theory (EMT)[61] is an approach to evolutionary theory that posits that the need to avoid certain costly errors will be reflected in the structure of the relevant entities, actors, or organizations the model is applied to. The motivating idea is that sensitivity to what possible errors exist and what the relative cost of different errors may be is something that needs to be reflected in the "design" of the relevant entities—it is an intuitively plausible principle of "good engineering" (Johnson, et al. 2013).

Take smoke detectors, for example.[62] The *raison d'etre* for smoke detectors is to alert us to fires—hopefully quickly enough to ensure that we can put out a fire before it becomes too great to manage or escape the building in case a fire cannot be managed. But most of the time when our smoke detectors go off, it is because they have detected smoke from cooking that we regard as unproblematic. If smoke detectors were designed to treat the false positives[63] and false negatives as equally problematic, they would be less effective in alerting us to fires and saving lives. They are biased to be overly sensitive, resulting in our smoke detectors going off more often than we would like, resulting in them being mostly

[61] The term was coined by Haselton and Buss (2000). In a way, this makes Haselton and Buss 2000 the point of origination for EMT. But error management considerations have been appealed to long before the publication of their work. Pascal's wager, for example, is an appeal to relative error cost. Cosmides and Tooby 1994, Higgens 1997, and Tetlock 1998 are all examples within evolutionary psychology that make appeals to relative cost of errors in their explanations. See Johnson et al 2013 for further exploration of places where EMT principles have been employed before and after Haselton and Buss's coinage.

[62] Note: the example of smoke detectors to demonstrate EMT principles was done first by Haselton, in writing that was on a now discontinued website (an archived version can be found linked on the Wikipedia page for Error Management Theory).

[63] False positives here are cases where we are alerted to a fire danger when there is none. False negatives are cases where there is a fire, but the smoke detector does not sound the alarm.

annoyances in our lives. But eliminating the annoyances they cause is not worth diminishing their effectiveness in preventing fires. We want them biased to false positives, because the false negative is the thing we really want to avoid. So a design choice that has zero false negatives and many false positives is an optimal choice for smoke alarms.

Application of error management strategies like this can be found wherever design occurs. Errors have different costs and treating them all as equally significant would amount to bad design.

In evolutionary theory, the lens of EMT can tell us something about how organisms have evolved, with mechanisms of natural selection favoring organisms that consistently avoid certain costly errors, even if the strategy for doing so comes with other costs.

It has long been held that nothing makes sense in biology except in light of evolutionary theory. More and more thinkers in cognitive science are extending this adage to the study of the mind. Nothing in the mind makes sense except in light of evolutionary theory.

Application of error-management theory as a means for discovering interesting structural features about human cognition has become more and more commonplace in recent years, with it being particularly prominent in the cognitive sciences of religion (Barrett 2004), conspiracy theories (van Prooijen and van Vugt 2018), and human mate selection psychology (Haselton and Buss 2000). While not everyone in the cognitive science wheelhouse is fully onboard with evolutionary psychology, and not everyone who is onboard with evolutionary psychology is adopting error-management theoretic frameworks, the case that error management strategies are an important factor in evolution seems strong enough that it may well become part and parcel of every cognitive scientist's toolkit in the future.

When applied to human cognition, error management theory is revelatory of biases in cognition, which shape belief formation away from strictly evidentialist norms. It can explain biases that we are already aware of and help us uncover biases we were not aware of, but which would make sense on an error-management-theoretic framework (McKay and Dennett 2009). When so biased, agents are more likely to reach conclusions and thereby form beliefs they are biased toward and less likely to form beliefs they are biased against.

Some of these biases are purported to be global[64] in their effect on cognition—though not necessarily universal, and certainly variable in their degree of influence on different minds. For example, in the cognitive sciences of religion and conspiracy thinking, a hyperactive agency detection device (HADD) has been posited to explain and predict religious thinking, by appeal to a bias to prefer agential and intentional explanations to non-agential ones (Barrett 2004, Lisdorf 2007).

Take for example a rabbit hearing a rustle in some nearby bushes. If the rustle in the bushes is just the wind, not much is lost by running from that rustle. If on the other hand, the rustle in the bushes is a predator, failing to run because one is in need of gathering further evidence or because the objective probabilities favor wind over predator leads to becoming lunch all too often.

Animals that are error avoidant thus fare better than animals that are more attuned to objective probabilities. Imagine, for example, that 99/100 bush rustles are just the wind, while only 1/100 bush rustles is a predator. Rabbit 1 believes there is a predator 100/100 times, while rabbit 2 believes is a bush rustle 100/100 times. Rabbit 2 will have 99% accuracy in its belief formation strategies, but will fail to avoid the costly error of getting

[64] What I mean by this is that they are a feature of how our minds are always processing information, if we have such a bias.

eaten. Rabbit 1 will have 1% accuracy in its beliefs, but will avoid being eaten. The result is that the rabbit with consistently false beliefs is viable, while the rabbit with consistently true beliefs is not viable. Rabbit 1 is much more likely to pass on its genes than rabbit 2.[65]

Other biases may be quite local, in their effects. Take for example the startling recent finding that hunger affects the likelihood of judges granting parole (Danziger et al 2011). The further away from a meal the judge is, the less likely the judge is to grant parole. From an error-management-theoretic perspective, this makes a lot of sense. Animals are more apt to secure their caloric needs by upping aggression and lowering empathy when they are hungry. If one must kill for food, then feeling bad for your lunch is a good way to miss out on lunch. It is no surprise then that people get cranky and combative when they are hungry. From an evolutionary error management perspective, that is a good response to have. Carnivorous animals that cannot kill to eat are apt to be weeded out by selection pressures. Herbivores that are not willing to muscle around their fellow herbivores to ensure that they get their share are apt to go hungry in times of food scarcity and be weeded out. Aggression and lack of empathy, even when not deserved, is a good error management solution for any species to arrive at in the face of hunger. But in the ecology of the criminal justice system, it results in an unequal application of the law, which is effectively invisible to the judges who are perpetrating it.

2 – The Error Management Analysis of Emotions

[65] I of course do not mean to suggest that either of these is an optimal strategy for dealing with bush rustles. I only mean here to produce an example that (a) demonstrates the importance of relative error cost, and (b) demonstrates that high accuracy of beliefs is less important than having the beliefs that avoid the costly error.

There is no consensus among philosophers as to what emotions are. One of the more prominent traditions in philosophy regards them as evaluations or judgments (Broad 1954, Kenny 1963, Solomon 1980, Neu 2000, Nussbaum 2001). Another tradition identifies our emotions with the distinct qualitative aspects of experiencing them, *the feeling tradition*—a view that seems to have been presumed by much of western thought, prior to the advent of modern psychology (James 1884, Lange 1885). Yet another tradition marries the previous two traditions, regarding them as feelings with intentionality (Goldie 2000, Helm 2009, Kriegel 2012). Another line of thought regards emotions as a distinct kind of perception (Roberts 2003, Prinz 2004, Tappolet 2016). Another identifies emotions with patterns of salience (de Sousa 1987, Ben-Ze'ev 2000, Evans 2001). Others take something about the motivational force associated with emotions to be what is essential (Deonna and Teroni 2012, 2015).

It seems hard to deny that in most things that spring to the minds of most people, when asked to think of an emotion, all of the above components are present. Typically, there is a judgment or evaluation, there is a distinct feeling, one's attention is patterned in a specific way, one becomes acutely aware of certain things, and one is motivated to do or not do certain things. What is at issue in the debate is which of these things is the emotion.

What also seems hard to deny is that there are cases where we are tempted to say we have an emotion, but where one or more of the above elements is missing (or is at least not obviously present). There are recalcitrant emotions (D'Arms and Jacobson 2003)—cases of our emotions persisting, despite our judgments not fitting our feelings, such as in the case of an irrational phobia or feeling scared when watching a movie, despite knowing there is no danger. There may also be cases where we are mistaken about what we are feeling, raising

questions about whether there is any distinct feeling associated with certain emotions. Lisa Feldman Barrett, for example, claims to have once thought she was falling for someone on a first date only to realize it was the flu (2017). There are also things we call an emotion, like love, which seem to require feeling different things at different times, like joy at the success of your love's object, or devastation at their funeral. So they are cases where there appears to be no distinct feeling associated with the emotion. There are also times where emotions leave us feeling paralyzed, robbing us of any motivational force. And there are times where our emotions seem to blind us in ways that seem very much unlike perception. So, there are putative counterexamples to all of the major theories.

Complicating matters further is the question of what counts as an emotion. We typically do not call things like hunger or thirst emotions, but it is not altogether clear how we can draw a principled distinction between them and other things we call emotions, like fear and anger. All of the properties that philosophers want to identify as the very thing that an emotion is seem to be present in the case of hunger and thirst.

There thus seems to be serious doubt about whether any one of the above accounts can match or vindicate our intuitive grasp of the emotions. For the purposes of this paper, it really is of no consequence which of these things, if any, are called emotion. Nothing in what I have to say here will require rejecting any of them. What is of consequence is that, depending on what one says an emotion is, there is likely going to need to be some rewording of my arguments and explanations, as what is constitutive of emotion, what causes an emotion, and what is caused by an emotion will vary according to one's preferred theory of emotion.

Emotions (and affective states more generally) seem ripe for an error management analysis. Affective states are operative in time sensitive decision making—often to an extreme degree. If one detects danger, the fear response is automatic. It patterns your attention toward the danger and toward finding solutions for extricating yourself from that danger. It causes you to ignore other valuable things in life. If running for your life, you are quite unlikely to be thinking about the exam you have in the morning, whether you will get a promotion, or about whether your partner may be cheating. (It is not unthinkable, just very unlikely.) And the fear response will cause you to be slower to trust, quicker to jump to further conclusions about danger, and more apt to be vigilant in monitoring their environment for danger—even if there is clear evidence the danger is over.

The need to be error avoidant in time-sensitive decision making and automatic responses seems an intuitively compelling use case for error management theory. Emotions also seem to be about the very things that error management theory is supposed to tell us useful things about. They are about the avoidance of costly errors, like threats that may provoke a fear or paranoia response, threats that may warrant aggression, and obtaining valuable species goods, like mating opportunities, calories, status, or security. Failure to attain those species goods or avoid life ending or permanently disfiguring dangers is exactly the kind of things that selection pressures will operate on.

So, an error management theory analysis of emotions seems likely to be productive. It would predict that affective states occur when a salient costly error is rated as being possible in one's environment. The affective response will pattern one's attention toward the source of error, retrieve from memory salient information for handling the situation, and bias belief formation away from the costly error—resulting in beliefs being formed on weaker evidence

than would normally be required or beliefs being held in the face of significant countermanding evidence, when that evidence would otherwise be sufficient to alter belief.

Let us take an example. If an agent believes they are being dealt with unfairly, they may have an anger response. This may lead them to adopt a combative attitude and ratchet up their aggression levels. They are likely to search for evidence within their memory that the person they are dealing with has a history of negative evaluations of or unfair behavior toward them. They are unlikely to retrieve from memory information that would speak against such conclusions. They are more apt to form beliefs that further events unfolding are unfair on weaker evidence than they otherwise would. If, for example, Julie has failed her qualifying exam in grad school, for reasons that she believes unfair,[66] we might expect a response from her that looks something like this: she will search her memory for instances that support a narrative that the faculty is biased against her, without searching for evidence to the contrary. She will comb her memory for reasons why the evaluation was unfair, without searching for evidence to the contrary. And she is more apt to take suggestions meant to address her concerns to be further injustices. She will be biased through and through to fight for her status that she is at risk of losing. She will be less likely to form beliefs about the good intentions of the faculty, less likely to attribute good character to them, and less likely to evaluate evidence in an unbiased manner. She will be systematically biased to belief formation and revision processes, habits of mind, patterns of attention, memory recall strategies, and behavioral repertoires that are apt to be preservative of her

[66] The wording here implies that the reason precedes the emotion, contra Haidt 2001. Nothing in this paper hangs on whether Haidt is wrong or right. So, the ordering here is best seen as a matter of having to pick some way of talking, rather than a genuine attempt to weigh in on the debate.

status and systematically biased against the ones that would concede the justness of her loss of status.

3 – Two-Tier Error Avoidance

Julie's story is a familiar one. We are never surprised to see someone think or act like Julie has when threatened with a loss of status. And we are never surprised when they start coming to some wild conclusions on a paucity of evidence or when they hold onto indefensible beliefs in the face of overwhelming contrary evidence. It is the very thing that has caused so many thinkers in the past to claim that the emotions make one irrational (more on this point later). This suggests that the error management theoretic analysis is capturing quite well important elements of our tacit understanding of the belief-emotion nexus.

But how can this make sense evolutionarily? Many philosophers take it to be a truism about belief that it aims at the truth (Williams 1973, Shah and Velleman 2005). And the idea that evolutionary considerations might speak in favor of such a conclusion is a common idea (Dennett 1971, Fodor 1983, Millikan 1984, Van Leeuwen 2018). Animals need true beliefs to successfully navigate the world. So, why would evolution favor belief formation mechanisms that systematically bias away from the truth?

The error management theory analysis has already told us why. But the concern does bring up an important point. When we deliberate with each other, we often presume that the strength of evidence is what should decide the merits of a belief. Intransigence in the face of evidence is often sufficient to license the inference that the person is being irrational. So, we are responsive to evidence in a way that seems oriented toward acquiring true beliefs. But

how to square that with belief formation mechanisms that are systematically biased to be error avoidant and not necessarily truth acquisitive?

One may be tempted to think that the answer is that we are only dealing with emotional responses sometimes and thus that the cumulative effect of such biases is negligible; thus, the overall system is truth-directed. But there is actually quite a bit to speak against that. A variety of affective phenomena, whose connections or distinctness are not altogether clear, are pretty much always active. Talk of *moods* for example is talk of an overarching glumness or cheeriness that colors vast periods of one's life. If one is in a foul mood, then one is apt to have a negative outlook on life—which seems tantamount to them patterning their attention toward the negative, forming beliefs with negative evaluations on lesser evidence, failing to search for evidence that would undermine their pessimistic expectations, and being resistant to forming positive expectations, even when faced with good evidence for them. A cheery mood may just well be the opposite.

Affective scientists speak often of *core affect,*[67] which may well be a characterization of mood in the absence of an emotional episode (the relation between mood and core affect is an issue for another day). The idea is that one's core affect is always operative, and when it swings toward negative valence or positive valence, one's whole outlook on life can change, leading one to be more apt to form certain beliefs and less likely to form others—more apt to engage in certain behaviors and avoid others too. It may well be that we are never free from core affect having some effects on our cognition. If this is right, then the idea that affective

[67] The term was coined by JA Russell (Russell and Barrett 1999 and Russell 2003). In Russell's view of emotion, core affect plays a role of being a consciously accessible "continuous assessment of one's current state" (2003), which is a component of every emotional experience/episode. Whether or not one accepts Russell's views on emotion, the idea that there is always a consciously accessible feeling that assesses one's current state, regardless of whether one is experiencing an emotional episode has proved to be widely popular.

states only bias our belief formation *sometimes* seems unlikely to be defensible. Affective states shaping our belief formation and revision practices is not an exception to the rule; it is the rule.

The alternative reconciliation I propose is a two-tier error avoidance system. This can be understood on the two systems model of cognition, with the first tier being the intuitive and affective *fast thinking,* while the second tier is the cognitively expensive *slow thinking* (Kahneman 2011). But if one has doubts about the metaphysics of these two systems (e.g. Zimmerman 2018), it can simply be understood as a two-step algorithm for error avoidance, which could be instantiated in a number of different ways on a number of different models of human cognition.

The first tier of the error avoidance system thus is comprised by a swift first pass that is affectively guided away from fitness undermining costly errors, even if doing so results in false beliefs or mistaken behavior.[68] But since our belief formation mechanisms are systematically biased across the board due to our affective states, we are at great risk for *error pileup* and *error entrenchment.*

Entrenchment and pileup are very much related, but they are distinct. Pileup is meant to be a measure of quantity of errors, while entrenchment is meant to be a measure of the integration of those errors into the cognitive system. The two are distinct but related. Errors may pileup without entrenchment, when beliefs are unimportant. If I am in a foul mood and mistakenly judge a cashier I will never see again as a bit rude, this error is unlikely to have much significance to my cognitive system. If I am paranoid and am convinced that someone

[68] In saying that it is error avoidant, I am not suggesting that any consciously accessible reasoning processes are scrutinizing the matter, critically assessing potential for error. Rather, unconscious information processing is involved, delivering a consciously accessible intuitive judgment that is likely to be believed.

is lying to me about a piece of trivia from a fantasy novel, my false belief about fantasy novel trivia is unlikely to have any impact. Many such errors pose no threat to the organism. But when mistaken beliefs are formed and they become the basis for inferences used to form new beliefs, they are taken into account in planning and responding to behavior, when the agent has gone on record publicly about them (and would lose face to recant them), or when they become integral to our daily activities, the need to hold onto that belief is raised and the cost of losing it is high. Abandoning an entrenched belief requires heavy revisions to it the agent's cognitive system. When people who were highly religious, for example, walk away from religion, it roots out a wide range of their behavioral and cognitive repertoire, robbing them of familiar habits, emotional regulation strategies, and connections, impairing their decision making for some time to come.

When entrenchment and pileup coincide, agents are apt to have a very distorted view of the world. A distorted view of the world can cause one to misdiagnose the salience of a costly error, which things would be a costly error, and when a species good is at stake.

Let's make that concrete with an example. A white male who mistakenly believes that white males are the most persecuted minority in the United States is apt to think that they need to act to avoid further loss of status. They might see acts aimed at remedying injustices faced by underserved communities or victims of prejudice as further persecution of their "minority." They thus think they are at risk of a costly error that is tantamount to loss of species goods like status and security, despite that not being the case at all. They are apt to systematically misdiagnose what is of value and what is at risk.

There are two reasons why this example is useful. The first is that it demonstrates how error pileup and entrenchment can distort one's view of reality, causing one to make costly

errors and try to avoid things that are not costly errors. But it also shows how, in certain ecologies, pileup and entrenchment may not always diminish fitness. Fighting to regain lost status, even if the loss is only imagined, may result in a privileged or advantageous position.

So, entrenchment and pileup are do not always diminish fitness. Traits that are useful in one ecology may be useful in another. And a trait may also vacillate between evolutionarily fitness promoting or diminishing within a single ecology. But if they cause one to miss or make costly errors, they can be disastrous. So, from an evolutionary perspective, it makes sense to think that systematically biased error-avoidant thinking cannot be the whole story, since the error pileup and entrenchment that would result would not make us very error avoidant in the long run.

A second tier of evaluation is necessary to root out errors resulting from the first tier, in order to counterbalance pileup and entrenchment and to remain error avoidant across time. Deliberation, reflection, and human interaction help to reduce problematic error pileup and prevent entrenchment. When we cool our heads and downregulate, we can think about the decisions we made and the beliefs we came to and check our work. Even with a cooler head, we should not think we are free from affective biases. But they are greatly reduced. And if we have company, we can talk to them about the beliefs formed and actions taken, in hopes that their insights will countermand any shortsightedness due to our own biases and cognitive limitations.[69]

Thus, with two tier error-avoidance, we are able to avoid costly errors in the short run and obtain species goods with a higher degree of probability than would be secured by mere

[69] It may well be that something stronger is true—namely that tier two is usually only ever engaged as a result of being prompted to engage in reasoning (Haidt 2001) or that tier two processes are inherently socially oriented (Sperber and Mercier 2017). Whether such ideas can be squared with what I have to say here is a project for another day.

reasoning by appeal to evidence. But we have a corrective for error pileup and entrenchment, so that we are error avoidant across time.

It is however worth flagging one last thing about this point. Since tier two processes are often calorically and cognitively demanding, it seems likely that we are not always running through the second tier. What seems likely to be the case from an evolutionary perspective is that evolution has fine-tuned us to engage in the minimal amount of second tier error avoidance needed to be ecologically viable. Let us call that point the *tier two equilibrium*—the place where, beyond that point, caloric needs were too great to be ecologically viable in our evolutionary history, and below which error avoidance was not sufficient to be reproductively successful in the long term.

My thesis then is that neurotypical human cognition is apt to congregate around this tier two equilibrium, with respect to engaging tier two error avoidance measures.[70] If that is right, we should think that much of our biased thinking has and will go unchecked.

4 – How what you believe depends on what you feel

I outline now two theses about how what you believe depends on your affective states. One of them has been an obvious presence already, but the other has been tacit and is apt to be surprising once made explicit. But some intuitive cases will support the claim that both theses capture assumptions that are already operative in our intuitive reasoning about the belief-emotion nexus.

[70] This might be achieved through the mean being the two tier equilibrium or it might be achieved through some evolutionary complementarity, wherein having a variety of deviations from the mean in either direction is evolutionarily advantageous for groups of humans.

The Biasing Thesis: belief formation and revision are systematically biased with respect to beliefs that are or would be valuable in avoiding a costly error, such as death or damage to the agent or missing out on a species good, so long as the possibility of such a costly error has successfully been detected and an affective routine occurred in response to it.[71]

The dependence relation the biasing thesis describes is not one where evidence has no role to play in belief formation. Belief formation, retention, and revision status for beliefs or potential beliefs that are not relevant to a salient costly error will not be biased. In the case of Julie, for example, she is unlikely to have biases that affect her belief formation with respect to questions about whether there is a window behind the faculty member she is meeting with or whether there is a notebook on the desk. A great deal of belief formation will be unaffected.

But belief formation and revision processes that are relevant to the avoidance of costly errors will be systematically biased. This bias is apt to occur in at least two ways. One of them is *motivated reasoning*, wherein the agent searches for evidence for a conclusion, but avoids as much as possible entertaining evidence to the contrary, no matter how strong it is. The other way is that one's intuitive plausibility judgments will be shaped in an error avoidant way, to ensure that the costly mistake is not made. In Julie's case, she will be more biased to form belief about the unfairness of what is happening, so long as she stays outraged, and less likely to form beliefs about the decency of the faculty members or the fairness of their evaluation, even if the evidence is strong. Until Julie downregulates, she is

[71] In saying that affective routines will bias us toward the avoidance of costly errors, I am not suggesting that there are not other sources of bias. I am also open to the possibility that there may be other species good which we have adapted to, such as preservation of mental health through preservation of positive self-image (eg. Mandelbaum 2018). There is room to expand on what other biases exist and what other costly errors may factor into our evolutionary history. But I regard such expansions to be complicated enough matters to be outside the scope of what can be accomplished here.

going to be extremely biased in her reasoning on the subject, and the longer she stays in that state, the more error prone she is going to be with respect to the truth.

I now turn to the second thesis, which I flag in advance is apt to be the more controversial of the two. I will also flag that I will be using the term *credence.* There is some question about whether credences are actually part of the furniture of the mind, and if so, how it is that they relate to degrees of belief,[72] outright belief,[73] or a belief with content that includes a proposition in the scope of a probability operator.

Nothing here really depends on credences actually being the correct entity. Rather, I have used the term because discourse around credences is apt to bring up associations with exactly the kind of stuff I want to point to. So, I suspect it is just the most readable word choice. But this work should not be understood as being wed to a theory of credences.

The Credence Shifting Thesis: changes in affective states can change the degree of credence given to a given proposition, without change of evidence. The result is that changing moods can change a mental state from a belief to a doubt, just by changing affective states—no steps in reasoning needed.

I suspect this one raises some eyebrows. But I think it is not too hard to find cases where the Credence Shifting Thesis is already operative in our understanding of the belief-emotion nexus. Consider Becky, who has a crush on Morgan. One moment, Becky is sure that Morgan loves her. The next, she is sure that he hates or disregards her. The emotional turmoil changes her attitudes in a flash.

[72] See Moon 2017 and Jackson 2019 for exploration.
[73] See Wedgwood 2012 and Hawthorne, Rothschild and Spectre 2016 for exploration.

Now, it may be objected that Becky is actually changing credences here as a result of new evidence. She may have suddenly thought of something that Morgan did, which she now suddenly takes as evidence that he has no regard for her, shifting the balance of evidence in a new direction. I fully acknowledge that this can happen. But t I am also convinced that new evidence is not needed for credences to shift. Becky can wake up one morning, feeling good, and sure that love is about to enter into her life and then get stuck in traffic, feel frustrated, and without even a single other consideration or change in evidence let her insecurity lead her to exclaim in full confidence "it is never gonna happen!"

If the credence shifting thesis is true, it suggests that we sometimes move away from true beliefs when doing so is error avoidant, but in an underhanded way, still keep the state around to be a belief again when the relevant error is no longer salient. A person may believe that, on the whole, Greg is a good guy. But when they find out that Greg is being consider for a promotion that they consider themselves deserving of, they may suddenly, in a flash, believe that Greg is a jerk and point out all the jerk-like things Greg has done in the past, working under the conviction that Greg is awful. And then, afterward, once they downregulate, they may send an email apologizing for having done Greg so dirty in the meeting—recognizing once again that Greg is an all-around good guy, and the evidence cited was overblown. In such cases, it appears that a change in emotion can suffice to change beliefs. This person may be willing to infer and plan with the attitude that Greg is a jerk and act on it as needed, only to return to the realization that this is not so, without any change in evidence.

5 – EB-Cycling

I have so far advanced an error management theoretic analysis of emotions; posited a two-tier error avoidance structure, which seems to correspond nicely with our actual belief formation and revision practices; and advanced two theses about the dependence of what we believe on what we feel. The picture on offer is not one where what we believe is *wholly* dependent on what we feel. But it is one where our affective states are a difference maker, which constantly biases us away from the truth in an error avoidant fashion, when the possibility of costly errors are made salient. Since not all beliefs pertain to a salient costly error and we have two-tier error avoidance, we are able to stay error avoidant across time—the second tier operating as something like an epistemic immune system, wherein incoming beliefs must be reconciled with existing beliefs and must survive the gauntlet of more cool-headed scrutiny.[74]

I now turn to cases wherein failure to downregulate and effectively employ two-tier error avoidance results in error pileup and entrenchment, and a phenomenon I call *Emotion-Belief Cycling* (EB-Cycling for short).

EB-Cycling occurs when the interplay between belief and emotion exhibits a circular, self-reinforcing structure, wherein emotions make the agent prone to beliefs of a certain kind; those beliefs in turn warrant continuation of the affective experience or more instances of it, which in turn causes the agent to form more beliefs of a certain kind, which makes them more prone that emotion, which makes them more prone to those kinds of beliefs, and so on.

[74] Mandelbaum 2018 argues for a psychological immune system, where belief formation and revision happens in a way that promotes good psychological health. It may well be that even at tier two, this psychological immune system still biases belief formation and revision practices.

Let's get a concrete example. Suppose Walter experiences an acute feeling of paranoia. He becomes convinced that his colleagues are talking bad about him behind his back, on the flimsiest of evidence. Enraged by the sleight and motivated to get to the bottom of it, he continues thinking about what this could mean, and finds some other puzzling social phenomena in searching through his memory—which in isolation would not be good evidence of negative attitudes against him, but in light of his newly formed belief that his colleagues are badmouthing him, seems like further evidence of negative opinions of his colleagues. He takes this as evidence of further hidden communications about him, sharing and spreading negative evaluations, damaging his status and prospects. This prompts Walter to see what else he was missing, finding more instances that, alone would not be good evidence of backbiting. But when taken in the context of his two newly formed beliefs acting as evidence and his feeling of paranoia making him more receptive, now seem to look an awful lot like evidence for a secret conspiracy of hate and badmouthing. Walter's paranoia and newly formed beliefs send him into a state of hypervigilance around his colleagues, wherein every social ambiguity takes on the force of confirming evidence for this hidden world of backstabbers, only deepening his paranoia. Walter's guarded and cold nature as a result of this only further prompt uncomfortable looks and silences, prompting more evidence for hidden negative evaluations. The paranoia makes paranoid beliefs come easier and more often and those beliefs warrant further paranoia. And so on.

Walter has begun to spiral, due to the self-reinforcing structure of how his beliefs and emotions are interacting. If Walter does not downregulate soon, he will be likely to cause the very negative opinions is trying to remedy. He may work himself into such a state of paranoia and vigilance that he will need psychological treatment.

Walter is EB-Cycling. His EB-Cycling is particularly worrying for his mental health and life quality. But it is also important to note that EB-Cycling does not always have such disastrous consequences. We might imagine, in Jamesian fashion, two shy people Yuki and George, who are so shy they never give any indication of interest in one another, but are secretly each in love with the other. Their feelings of love may cause them to believe the other is interested, despite a paucity of evidence. This may cause them to interpret any and every signal as a sign of love, even if it was not so intended. These "signals" may then be drawn on as evidence warranting deeper feelings, more belief formation on weak evidence, and so on. It may well be that Yuki and George EB-Cycle their way into a once-in-a-lifetime romance this way.

So, given that emotions are aimed at avoiding costly errors, such as missing out on species goods, it should come as no surprise that they can cause one to secure species goods in cases where mere reliance on evidence and an even-minded evaluation of its merits would not procure that good. Yuki and George are apt to look back at the situation and marvel at how crazy they were. They might describe the situation as one of not letting doubt get in the way and just going for it. Emotions do steer us toward valuable goods and away from costly ills. And when we are lucky enough to be in a situation wherein deviation from evidential norms is exactly what we needed to procure a good, we are apt to marvel at the power of emotions to so effectively steer us. We are apt to celebrate their hidden wisdom.

So, not all is doom and gloom for EB-Cycling. But there is certainly some doom and gloom to be had. It is easy to focus on the cases where listening to our emotions procured for us a good and marvel at the hidden wisdom of emotions. But we must also remember that it does not always go that way. Sometimes listening to one's emotions can lead us to EB-

Cycle in a way that leads us very far astray from value and on a collision course with disaster.

In the case of Walter, he needs to downregulate to engage the second tier to avoid going astray. In the case of Yuki and George, downregulating and more fairly evaluating the evidence could have robbed them of the good. We can also imagine the case of Yuki and Georgina, wherein Georgina is convinced Yuki is a lesbian who is in love with her, on exactly the same level of evidence that George was working with, leading to Georgina EB-Cycling her way to behavior that is tantamount to stalking and which proves life-destroying for both Yuki and Georgina.

There simply is no general fact of the matter about whether the consequences of EB-Cycling are going to be good or bad. When it comes to love, it probably depends on how attractive you are, with more attractive people more likely to have their EB-Cycling rewarded and less attractive people more likely to have their EB-Cycling come off as positively creepy. When it comes to paranoia, Walter seems to be on a terrible course. However, the EB-Cycling of an agent that actually has hidden actors and intentions working against them may fare better.

But on the whole, EB-Cycling can compromise our error avoidance capabilities, by failing to consistently engage tier two, resulting in pileup and entrenchment. It is no surprise then that evolution has equipped us with emotion regulation capacities, which are often sufficient to prevent EB-Cycling. It is also no surprise that we collaboratively police people's emotional responses, gently soothing them to calm and downregulate when they have failed to self-regulate or reminding them how they tend to get carried away in fantasies

when they fall for someone. The collective impulse to do such policing helps prevent and mitigate the ill effects of EB-Cycling.

6 – EB-Cycling in polarization, radicalization, and believing badly

There has been no shortage of literature on the question of why people believe things that are weird, stupid, or vile.[75] It has gotten to the point wherein people working on the topic feel the need to preemptively acknowledge that *this is yet another work on why people believe weird things.* I confess, I do not share the exhaustion that such authors presume will be present in their audience when these matters are discussed. I am inclined to think that we still do not have very good answers to these questions. But I do think that EB-Cycling needs to be part of whatever story we have to tell about such things—though I doubt it could possibly be the whole story.

Following Neil Levy (2021), let us speak of cases of such belief formation, retention, and revision as cases of *believing badly*—that is, cases where one is doing a bad at the job of being an effective believer. Levy argues that the cause of believing badly is ecological—it is the information environment we are in. Levy thus urges caution in evaluating bad believers, claiming that it is unreasonable to have expected them to do better in their present ecology, filled as it is with misinformation. Their bad beliefs are rational responses to their environment and the evidence they have.

While I am sympathetic to Levy's points, I am inclined to think that his analysis is missing many personal-level causes of bad beliefs. Levy is convinced that believing badly is almost entirely a matter of our bad belief formation environment. I am inclined to think that

[75] Note: include sample citations

believing badly is, in large measure, a product of *feeling badly.* By *feeling badly*, I do not mean to say that one is experiencing negative affect. Rather, I mean to say that one is failing to manage one's affective life effectively enough, resulting in negative impacts on cognition and belief. Where Levy thinks the problem is that people are being lied to, I think the problem is that they are being made angry and then lied to, making it so they are more apt to accept lies as truth, and more likely to get angry about the given information, thus becoming more and more susceptible to more bad beliefs, and EB-Cycling their way to some astoundingly bad beliefs.

According to the account I have presented, bad believers are also *bad feelers*. Failures to emotionally regulate often result in error pileup and entrenchment, making them more prone to EB-Cycling, and forming some truly astounding beliefs on the flimsiest of evidence. The further entrenched such errors are, the harder they are to extricate.

EB-Cycling is a danger. Error pileup and entrenchment can easily cause people to misevaluate moral stakes. EB- cycling can result in a disconnect from reality that can disguise moral obligations, making the obligatory seem wrong or merely optional, and making evil acts seem obligatory. People riled up over imagined injustices commit murders and genocides, go to war in defense of evils, hold up pizza parlors, and do all manner of terrible things. There is a reason why public speaking is often impassioned. It can move belief formation and action decision criteria away from what the evidence supports.

I flag once again that emotions' influence on our belief formation, retention, and revision practices is not always a bad thing. Yes, demagogues may utilize emotionial contagion to infect us with emotions that shape our belief formation in a way that makes us susceptible to being dragged into war and injustices, with disingenuous appeals to patriotism and justice.

But inspirational leaders may also motivate us feel compassion and alleviate injustices, we would otherwise ignore. In the war for the general populous' beliefs, emotional appeals have been used to steer us toward goods and away from ills, but the error avoidant tendencies of emotions have also been capitalized upon to entrench racist and sexist attitudes, wage preemptive and unjust wars, and spread hate. The collective debate about how to conceptualize the world is constantly being waged on emotional grounds, which seek to correct the limitations of evidence-based decision making in a populous that is inadequately informed and cannot devote adequate attention to becoming informed, and who are not all cognitively adept enough to reason critically and carefully about the issues.

But if we want to understand dangerous political polarization and radicalization, we have to understand the belief-emotion nexus. We have to understand how emotions bias our thinking away from costly errors in short-term decision making, and how successful and frequent engagement of tier two error avoidance to mitigate pileup and entrenchment is important to avoiding negative outcomes of polarization and radicalization. Anyone who has tried to talk to a person with outlandish political beliefs knows that one of the first things they will try to do is make it emotional. Their impassioned appeal seeks to infect you via emotional contagion so that your reasoning will be systematically biased in a way that will make you more receptive to their points.

If we want to understand how people of all intelligence levels end up joining cults,[76] joining fringe movements, leaving their first-world comforts to join ISIS, bombing abortion clinics, and believing and spreading conspiracy theories, we would do well to attend to how they felt along the way. It is an empirical prediction of this essay that along the journeys of

[76] For exploration of education and intelligence levels in cult members, see Dawson 1998, Stein 2017, and Rousselet et al 2017.

such people, we will find an emotional journey, the documentation of which will include emotional experiences that explain the ways such agents' belief formation and revision practices were biased to be receptive to certain beliefs and considerations and resistant to others. Staying in those emotions or moving onto new ones, without checking their work, results in pileup and entrenchment, skewing their intuitive plausibility judgments, which in turn inform their assessments about who is a trustworthy source of information, and what should be believed on what degree of evidence. I predict that we will find that people who are *bad feelers*, even if they are otherwise very intelligent, are most often among those who are *bad believers.*

I, of course, do not wish to say that EB-Cycling is a complete story of bad believing, radicalization, or polarization. I mean only to say that it is an important part of the story. I suspect it may be one of the most important parts of the story.

7 – Emotion and Rationality

It may be thought that this essay is a return to a traditional view about emotion, which has lately had a war raged against it. That traditional view is one on which reason and emotion are at war—a view wherein emotion is an unruly horse that has to be reigned in by the charioteer of reason.

The way I see it, however, is that the present view strikes a middle position between those who wish to celebrate the rationality of the emotions and those who wish to see them as being in tension with rationality. Let us call the new wave of thinkers, who celebrate the rationality of the emotions *the optimists.* And let us call those who are pessimistic about the rational impact of emotions *the pessimists.*

I am tempted to call my position the *realist* position, since it strikes a middle position, but that is kind of an unfair framing of the debate. *Realist* suggests that I am the one who has gotten it right, when obviously each side thinks they are the one who is in touch with the reality of the situation. So, let us instead call my position *the ambivalent position.*

The ambivalent position is that emotions are undoubtedly ecologically rational (Gigerenzer 2008). Integration of emotions into one's life is undoubtedly of value. We would not survive at all without our emotions (Damasio 2004). And so I do not think that we should seek to live lives of dampened emotion, always mistrustful of emotion, and celebratory of the efficacy of pure, cold-hearted reason. I cannot help but think that reason is overrated by such people.

The ambivalent position is also that emotions often cause us to deviate from epistemic and practical rationality in a way that is error avoidant. From an evolutionary perspective, this is a good thing—it is fitness promoting. A creature who always avoids the relevant costly errors, but has a lot of false beliefs is doing better than a creature who has a great number of true beliefs, but ends up in another animal's stomach. From an evolutionary perspective, true beliefs are valuable only to the extent that they are error avoidant.[77] And they are often error avoidant. But what one is most justified to believe is not always true,

[77] An objection raised here by an Aaron Zimmerman is that animals often take calculated risks, which may put their life on the line. This shows that they are not always error avoidant and seem to require being attuned to an objective probability assessment, which in turn requires true beliefs. But this is not at odds with anything I am saying here. Missing out on species goods like food or status is costly. So, never incurring any risk to life is not necessarily error avoidant. A predator that never puts itself at risk of death in pursuit of a prey is not a viable predator. We should expect some risk tolerance when the reward for incurring risk is great. And we should expect animals that cannot avoid entrenchment and pileup to be worse risk assessors. Saying that evolution has attuned our cognition to be error-avoidant with respect to certain costly errors is not to say that there are no circumstances where risk of error is to be incurred. There is a larger discussion to be had here about how probability assessments and risk taking fit into the evolutionary story, which I cannot address here. But suffice it to say, risk-taking is sometimes exactly what an EMT account would predict, and learned cognitive strategies and cognitive behaviors may often be in conflict with evolved cognitive strategies and behaviors.

and so deviating from standards of justification and defensibility in an error avoidant fashion is a better survival strategy—at least in the short term. Pileup and entrenchment may make an agent unsuccessful in error avoidance in the long run.

The ambivalent picture is that ecological rationality is most predictive of human behavior and fitness, but that practical and epistemic rationality are of instrumental importance to that end. They are compromised only to the extent that compromising them is integrated into a system that is robustly error avoidant in both the long and short term.

But what has been important for our survival and reproduction in past ecologies is not always a recipe for success in our present ecology. It is also often not a good recipe for justice. In our present ecology, the effect of emotions to cause deviations from epistemic rationality often prove to effect cases of doxastic wronging, social injustices, and both systemic and interpersonal racism, classism, sexism, homophobia, transphobia, ableism, and neuroprejudice.

The ambivalent position is not the simple, good or bad judgment, or the rational or irrational judgment about emotions. Rather, it is the position that emotions have a range of positive and negative effects on our lives and communities, and have a range of positive and negative effects on rationality, which may be good or bad depending on the situation. Recognizing the ways emotions can distort our view of reality and lead us away from good ends, hide from us the truth and make invisible our own biases and irrationality, is of crucial import for knowing when to self-regulate, when to assist others in emotional regulation, and when to let the emotions out so as to have their good effects.

So, I regard the questions of whether emotions are good or bad, or whether they are rational or irrational, to be mostly confused. We cannot live without emotions, so we have to

figure out how to live with them, mitigating their negative effects and cashing in on their positive ones. We have to recognize that emotional appeals may lead us away from epistemic rationality, but we also have to recognize that there are good times for that to happen and bad times for it to happen. Reason and reflex are not enough; we need emotion. But we cannot be uncritical in our approach to emotion. We need to check their work. But this does not mean that reason needs to run the show. It means that finding a healthy interplay of emotions and reason is important and that constant calibration of that interplay is needed to be adaptable to the wide range of circumstances that we face in life. Sometimes we need to stop thinking and feel. Sometimes we need to stop feeling to check our emotion's work.

When dealing with others, we have to recognize that emotional contagion can both be a risk and an asset. The emotions of those who are well regulated, who are responsive to value, and who are careful in forming beliefs are of immense value. When they infect us with their emotions, we are apt to be biased away from bad judgments and toward true beliefs in a way that may outstrip our own abilities or evidence base. But when we are infected by the emotions of those who are not well-regulated (the *bad feelers*), who are not responsive to what is actually of value, or who are careless in their belief formation, we are apt to be led astray from the truth and from good outcomes. The emotions of others, when they are good feelers, are a means for bootstrapping your way to good conclusions and outcomes. They are an asset. Being surrounded by people who are good feelers and good believers is like gaining a superpower. They will consistently bias your thinking toward good ends, in a way that your own capacities could not.

But when you are surrounded by people who are bad feelers and bad believers, their emotional contagion is apt to be disastrous. They will bias your thinking away from valuable ends, resulting in a disguising of your moral obligations and obscuring the moral stakes. Their error avoidant systems are likely to have gone awry, EB-Cycling their way to disastrous ends.

Being good feelers thus requires being epistemically careful, but being good believers often requires being good feelers. Emotions are not rational or irrational. Shutting them out will not succeed in making anyone rational. Living emotionally vibrant lives will not necessarily make anyone less rational. Having reason reign in the emotions is sometimes valuable and sometimes not. Different ecologies and circumstances may require different levels of emotional reactivity or regulation. There just is no simple story about emotion and rationality. There is no soundbite to be had about the rationality of the emotions.

Appreciation for the belief-emotion nexus, I believe, should lead us to the ambivalent position—the position that cashing in on the value that emotions bring and avoiding the disasters they can lead us to requires care and constant tuning. Appreciation for how emotions can affect our belief formation, retention, shifting, and revision is an important piece of the puzzle for figuring out how best to tune them to our ecologies and circumstances.

8 – Conclusion

I have attempted here to theorize the belief-emotion nexus. I have provided an error-management-theoretic analysis of emotion, which takes emotions as being responsive to detection of the possibility of a salient costly errors and biasing our belief formation,

revision, and retention in a way that guides us away from costly errors, even at the expense of acquiring beliefs that are false or are accepted or rejected in a way that deviates from evidentialist norms. I have posited a two-tier error avoidance structure that allows us to be biased in our thinking in the short term, to the avoidance of costly errors in time-sensitive decision making and belief formation, but which avoids the problems of pileup and entrenchment that comes from such biased thinking, by checking the work of the emotions, utilizing more calorically demanding reasoning processes and the collective brainpower of social relations.

I have used that framework to shed light on a number of puzzling issues that afflict us in our present social ecology and that have been debated by philosophers. I have made the case that we are prone to EB-Cycling, where a self-reinforcing structure of biased belief formation resulting from an emotion can justify further experiences of that emotion, which in turn lead to more biased belief formation, creating high risk for pileup and entrenchment.

I utilized this framework to cast light on the phenomenon of bad believers, such as those who hold politically extremist beliefs or who have been radicalized. I also weighed in on the question of what this means for the age-old debate between optimists and pessimists about the rationality of emotions, making the case that emotions effects are mixed and that the questions are mostly confused. Emotions are sometimes great shortcuts to positive ends. But they are also often shortcuts to disastrous ends. Being a good feeler and a good believer are intertwined to the point that it is hard to be one without the other. But being bad believers is often a result of the disastrous effects of being a bad feeler. Thus, emotion has both the power to improve and distort our reasoning. And its distortions can have both positive and negative effects.

While I have little doubt that there is much more that philosophers on all sides of these issues will have to say about my arguments here, what I hope this essay accomplishes more than anything is recognition of the importance of the belief-emotion nexus. It is something that cannot be ignored.

V. Believing the Victim

My use of *believing the victim* here is meant to pick out a family of slogans, hashtags, campaigns, movement, and policies. Such movements often use "believe the victim" as a slogan or may use some variant like, "start by believing," "believe women," or "believe survivors." Such movements have often arisen in response to a perceived moral failure of not giving sufficient credence to the testimony of those who claim to be victims of injustices. Such movements have arisen in response to the treatment of reports of sexual assault and of the treatment of those making the reports. And it is the issue of sexual assault reports that will be the focus of this paper. But such language has also begun to be used more recently in support of victims of racism, misogyny, transphobia, and homophobia. Often it is the case that those who do not suffer from these injustices are blind to just how prevalent these injustices still are, and they often dismiss the testimony of those who have been victimized, thereby depriving the victim of validation, support, and the justice, protection, and peace of mind that would come from imposing a corrective on the perpetrator's behavior. While some of what I have to say here will be generalizable to these other cases, not all of it will be.

Before I dig deeper into the focus of my essay, let me pause a moment on the use of the word *failure* in the above paragraph. There is a thin sense and thick sense in which we might want to use the word *failure*. In the thin sense, *failure* here just means not living up to some appropriate standard. It carries no claim about the culpability of anyone involved. People may fail to live up to appropriate standards for reasons that are not their own fault.[78] So, to

[78] Crewes and Ichikawa 2021, for example, expresses sympathy with the idea that the problems of not taking the testimony of complainants seriously enough may be a systemic and structural problem, not a problem at the level of individuals withholding belief. They take themselves to be adopting a Frickerian

say that believe the victim movements are responses to a moral failure in this thin sense need only mean that we are not living up to a just standard in the way we deal with those reporting injustices. It need not carry any claims about whether anyone responding to reports of injustice is or is not culpable for their failure to believe victims. In the thick sense, *failure* denotes not only not meeting the correct standard, but doing so culpably. People claiming there is a moral failure to give appropriate credence to the reports of those claiming to have suffered injustice may well be convinced that there is a moral failure in both the thick and thin sense. But for the purposes of this paper, I only intend to use *failure* in the thin sense, so that I can narrow the focus of this paper down to a manageable chunk of the overall problem.[79]

So, movements built around believing the victim are responses to the perceived moral failure of not giving sufficient credence to reports of injustices suffered. While the issues here extend beyond reports of sexual assault, I'll be focused here on only reports of sexual assault. Further, while those interested in proclaiming "believe the victim" are not only interested in the way sexual assault reports are treated in institutions, like universities and police departments investigating the reports, I will, for the sake of keeping the discussion focused, talk only about the institutional investigations of sexual assault reports and of claims that institutions are failing to give appropriate credence to these reports. Issues about

approach to the issue (Fricker 2007). Mason 2021 explores the issue of sexist ideologies in cases of sexual assault and the moral responsibility issues that come with it.

[79] For exploration of the issue of how we might make sense of the blame and praise issues associated with cases of the thin sense, where there is a temptation to blame but no source of culpability, Mason 201p provides a useful overview of the topic and a provocative and influential take on the issues. For exploration of the issue of responsibility for belief in particular, Owens 2000 and Zimmerman 2018 offer some worthwhile discussions of the matter.

legal standards, criminal prosecution, or public attitudes at large will mostly be pushed to the side.

Now, there are a number of challenges to believe the victim movements, campaigns and policies in both popular and academic discourse. Some of them are quite obviously bad faith contributions to the debate, motivated by misogyny, rape culture, or long-debunked myths about the extent of false reporting. And some criticisms are not criticisms of "believing the victim" *per se*, but rather of clumsy implementations of policy, such as interviewing techniques meant to champion the cause of victims, but which run the risk of distorting memories or implanting false ones (Davis and Loftus 2019). No doubt, such issues paint the believe the victim movements in a bad light. But they are not objections to the core idea of believing the victim—namely, that there is a moral failing that needs to be addressed, in order to see justice done for victims. These challenges will also be pushed to the side.

The challenges I am interested in engaging with are the ones having to do with core philosophical issues, which connect up with the nature and ethics of belief. As I see it, there are at least three of these, pertaining to the rationality, agency, and value theory questions with which this issue is entangled. This paper is concerned with *the rationality challenge*—the question of whether any changes in policy aimed at getting victims more reliably believed could be rational. The issue of whether it even makes sense to adopt policies aimed at believing victims, given that belief is largely[80] involuntary (I call this *the coherency*

[80] I hedge here with "largely," as those defending some form of doxastic voluntarism tend to regard doxastic control as being limited in scope. There is perhaps no truly uncontroversial adjective I could pick here, as it is difficult to quantify beliefs, and if beliefs are infinite in quantity (chapter 1), then it may be issues with quantification of proportion of beliefs with any properties, since it is conceivable that there may be infinite quantities of voluntarily held and involuntarily held beliefs, if some form of doxastic voluntarism is true. See Williams 1973 for the classic argument that belief is not voluntary, and Bennett 1990 for elucidation. See Zimmerman 2018 and McCormick 2014 for contemporary exploration of doxastic voluntarism.

challenge); and the issue of whether any such policy could be desirable, given that it seems to amount to institutional control of the beliefs of employees (I call this *the desirability challenge*); cannot be discussed here.

My aim here is to defend the rationality of believing the victim. But it is probably not very clear just what that means yet. Obviously, everyone believes, excepting perhaps the perpetrators, that victims should be believed. We do not have an issue of people denying victims should be believed. We have an issue of victims not actually being believed, because people are convinced that they are not actually victims. And so, it may be worried that the explanation for why many victims are not believed is because doing so would be irrational, given the amount of evidence available to investigators. Thus, it may be worried that chanting "believe the victim!" may amount to chanting "reject the standards of rationality, in order to side with the complainant!" Ensuring that victims are believed may amount to requiring believing without evidence, believing against the evidence, or as Debbie Moak, of the Arizona governor's office has claimed, introducing a confirmation bias into investigations of sexual assault reports that undermines the rationality and integrity of the investigation. So, that is the rationality challenge.

My aim here will be to answer the rationality challenge. I aim to make the case that (1) believing victims is rational more often than victims are believed; (2) that, all else being equal, believing complainant's testimony over the testimony of the accused is a rationally acceptable position for the start of inquiry; (3) that believing the victim policies do not introduce any investigation-undermining bias; (4) that even if they did introduce a bias into investigations, there is good reason to think that doing so would improve the accuracy and rationality of investigations; (5) that there are a number of policy changes that fit under the

heading of "believing the victim" that could *improve* the rationality of investigations, without compromising their integrity; (6) that the evidential value of complainant's testimony is often underestimated, and thus discounted on inadequate grounds; and (7) that the rationality of believing the victim is in no way dependent upon controversial theses like pragmatic or moral encroachment. If I am right about even a fraction of these points, the rationality challenge is severely undermined. If I am right about most or all of them, then it makes good sense to say that believing the victim is rational.

In section 1 of the paper, I will offer some initial considerations for a positive assessment of the rationality of believing victims. In section 2, I will consider challenges to the claim that believing the complainant in a sexual assault report should be the default position at the start of inquiry and show why they fail. In section 3, I will turn to questions about our obligations to victims after the start of inquiry. In section 4, I will consider Rima Basu's suggestion that believing the victim might be a case where appeal to moral encroachment is necessary and make the case that she has things backward. In section 5, I'll take a close look at common claims that who to believe in a sexual assault investigation is often a he-said-she-said standoff, with no reasons to favor the testimony of either party, and make the case that it is an indefensible view.

Section 1 – Licensing some optimism

Is it rational to believe complainants reporting sexual assault and should we think that we can adopt policies which do a better job of ensuring that victims are believed without compromising the rationality and integrity of investigation of sexual assault reports? I believe we should answer with a resounding "yes!" The aim of this section will be to make

an initial positive case that we have good reasons to believe complainants and that policies aimed at belief in victims can improve the rationality and accuracy of investigations. More details will be added to this positive case in following sections, where I address criticisms of believing the victim policies. But before we even get to the criticisms, I want to make an initial case for the view that rationality is not only not threatened but improved.

In 1.1, I will take a look at the statistical evidence and make an initial case that it supports believing complainants and doubting the accused, all else being equal. In 1.2, I'll turn to some considerations about evidence gathering and make an initial case that there is good reason to think that believing the victim policies are effective ways of improving the rationality of evidence gathering.

1.1 – The he-said-she-said standoff and the statistics

It is often thought that the dialectic of a rape case is something like a *he-said-she-said standoff*.[81] The complainant gives testimony that supports believing the accused is guilty. The accused gives testimony that absolves the accused of guilt. In the absence of any other telling evidence, the evidence is thus equal. And if the evidence is equal, it might be thought, then there is no case to be made either way. It is the word of the complainant

[81] A critic may wish to say that there is something objectionably heteronormative about this characterization, as it seems to presume that the only cases of interest are cases where of the complainant and accused one is always a man and the other is always a woman. But cases of sexual assault involving two members of the same gender or involving non-binary individuals also occur and are just as serious. While I see the worry, the locution of he-said-she-said is common enough to have heuristic value in getting readers onto the subject matter. But there is an issue about the ethics of discussing sexual assault that is looming here. Literature on the subject often is heteronormative and often proceeds as if issues of men raping women were the only issues. Defenders of these norm will be quick to point out that men raping women is the most common form of rape. Critics will be quick to retort that even so, cases of rape that do not adhere to these gender norms are common enough and psychologically damaging enough that it is important not to marginalize such victims. While I cannot adjudicate the debate here, I can proclaim that none of my arguments here require, nor are my conclusions limited to cases of sexual assault that have any particular gender characteristics. The choice of keeping "he-said-she-said" is merely for its heuristic value. It is not meant to be a statement about which rapes are worthy of discussion or attention.

against the word of the accused and each of their words is of equal evidential worth. But is this really the best way of looking at it? There are some good reasons for doubting it is.

If we just start from the statistics about rape cases, it seems as though at most ten percent of cases in the United States today involve a false accusation.[82] I say "at most" because experts on the issue believe there are some good reasons to think the number is much lower than that. People may recant their testimony, because of bullying, bribery, or sheer exhaustion from an investigation that is deleterious to their mental health. What's more, detectives may conclude that a victim is lying when they are not. Consider, for example, the famous case of Marie, whose own foster mother thought she was lying, because she had a moment of levity the day after being raped and because she wanted the same type of bedsheets that the detectives had taken away as evidence in their investigation. Marie's foster mother was convinced this was not how a rape victim would act and reported her suspicions to the detectives, who then exhausted and bullied Marie into making a false confession of lying.[83] While Marie's case is extreme, it is demonstrative of the kinds of things that those reporting sexual assault are often subject to.[84] While there is no consensus about an exact number, a number of research studies have concluded the number is lower, with research teams reporting the number to be 2% (Kelly et al 2005 and Heenan and Murray 2006), 3-4% (McMillan 2018), 4.5% (Spohn et al 2014), 5.9% (Lisak et al 2010),[85]

[82] See Lisak et al 2010 and Ferguson and Malouff 2016 for a wider look at the various estimates that researchers have come to.

[83] See Miller and Armstrong 2015 for Marie's story.

[84] Another complicating factor is that we do not know how many have been falsely accused, but were convicted—the falsity of their accusation never having come to light. Much of the recent exploration of the statistics has been interested in correcting widespread myths about the commonness of false reports, in order to get victim's testimony taken more seriously. The issue of how many may have been falsely accused and convicted seems both harder information to obtain and likely to be information that people are less motivated to uncover.

[85] See Lisak et al 2010 for an examination of the evidence and a defense of the 2-10% range and for an overview on the debates about where in that range the number should be placed.

or 7.1% (Lonsway 2009). Ferguson and Malouff's (2016) meta-analysis of various studies placed the number at 5%.

Whichever figure is correct, it poses a serious challenge for the he-said-she-said standoff view. In the case of sexual assault reports, all else being equal, it is highly unlikely that the report is false. The facts that (a) sexual assault is unfortunately all-too-common,[86] (b) victims seldom have any incentive for making a story up,[87] and (c) that false reports are rare are all compelling evidence speaking in favor of believing the victim. The most reasonable belief to have about why someone is reporting a sexual assault, all else being equal, is that they genuinely and sincerely believe they were assaulted.

Contrast this with the case of the accused. 90% of the accused are guilty. Rapists have a strong motive to lie: to avoid punishment. Rapists are typically more morally flexible than the average population. They have demonstrated a serious moral failing in violating the autonomy of their victims. They thus are less likely than your average person to have qualms about lying. There may be some rapists who draw the line at lying, but it is hard to imagine they are representative of the population. Further, while I don't have any statistical research to back this up, I suspect the number of criminals who lie about having committed a crime is pretty high. It is hard to believe that it could be as low as the amount of people that falsely report rape. And so, given that the accused is statistically unlikely to be falsely

[86] It is estimated that there are 433,000+ victims of sexual assault each year in the United States (Dept. of Justice 2020). As of 1998, it was found that one in every six women had been the victim of rape or attempted rape (National Institute of Justice 1998).

[87] The most common motives for false accusation are alibi, revenge, and attention. The first identification of these three as the most common motives for false reports occurred in a now much-maligned study, Kanin 1994. While much of what Kanin had to say in his work has been debunked, in this one point alone he has been vindicated by later work. McNamara, McDonald and Lawrence 2012 reports the same findings. De Zutter, Horselenberge, and van Koppen confirm that these are the most common motives, but also found that 20% reported that they did not know why they did it, and that there were also many cases of false reports that did not fit into any of these aforementioned categories.

accused and that most of those who are guilty and accused are liable to lie, it follows that most of those who are accused of sexual assault and deny having committed one are lying.

If that point is not obvious, we can do the math. Let us make a conservative estimate and say of the guilty accused, only 50% will lie about their deeds. I suspect the number is likely to be closer to 90%. But being more conservative will better suit our purposes here. And let us assume that the statistics of false reports is that 10% are false reports. (As I mentioned above, expert consensus suggests it is even lower.) This means in a sample of 100 sexual assault reports, 90 of the reports will be true. 10 of them will be false. Half of the 90 guilty accused will deny having committed an assault. So about 45 will say they are innocent and be lying, and the ten innocent people will say they are innocent as well, resulting in 55 people claiming innocence. This means that even if only half of the rapists lie, which is generously conservative, the chances that an accused person claiming innocence is lying are still over eighty percent, with 45 out of the 55 claiming innocence lying. If the rate at which rapists lie about having raped is higher or the number of people who falsely report sexual assault is lower, as I expect they both are, the likelihood of the accused testimony being false will be even higher. If rapists lied about their rapes at the rate of 90% and false reports were only 2% of the total number of cases, the likelihood of the accused's testimony of innocence being true would be a little over 2%. So, from the rareness of false sexual assault reports and the likelihood of lying rapists alone, it follows that the accused is most likely not telling the truth. Intuitions to the contrary seem to either be in ignorance of the statistics or to be guilty of ignoring base rate information entirely.[88]

[88] Aaron Zimmerman has posed a worry for this—namely, that it may overgeneralize. Perhaps it is the case that all crimes have statistics that will look like this. Does this mean that it is rational to believe any accusation of any crime? I see no problem in saying, "yes." It being rational to believe does not mean that there is enough evidence to meet standards of conviction. What is more, as evidence mounts, the rationality of

These considerations suggest the following lines of reason:

The Believing the Victim Argument
(BV1) It is statistically highly unlikely the complainant is lying
(BV2) If so, then, ceteris paribus, it is rational to believe the complainant[89]
(BV3) So, it is rational to believe the complainant

The Doubt the Accused Argument
(DD1) It is statistically highly likely the accused is lying
(DD2) If so, then, ceteris paribus, it is rational to doubt the accused
(DD3) So, it is rational to doubt the accused

Those who wish to defend the he-said-she-said standoff view will have to pick a premise to deny in these arguments. They will either have to deny the correctness of the statistics (first premises), or deny that statistics determine what it is rational to believe in cases where all else is equal (second premises).

A third strategy might be to accept the conclusions of these arguments, but insist that the standards of justice require ignoring the statistical evidence, in favor of evidence that is specific to the cases. This is an interesting issue and deserves a paper of its own. But for our purposes here, it need not be addressed. My concern here is with whether believing the victim is rational. And this is a response to the effect that justice requires being irrational,

---

believing is apt to change. What is at issue is whether we should believe the initial report of the complainant over the accused. And if the statistics warrant it, I see no reason to deny the rationality of starting out by believing. This of course does not mean that one is capable of believing. Nor does it mean that we should convict everyone we believe guilty. It just means that, in the absence of defeaters, the person conducting the investigation who believes the complainant over the accused, in cases where the statistics warrant it, is not being irrational.

[89] Perhaps the easiest way to get a handle on this ceteris paribus clause is to start with a case wherein the only evidence we have is the testimony of the complainant alleging the assault and the testimony of the accused denying it. Neither side has any obvious defects in their testimony that should cause us to doubt it. In this case, it seems as though we have no other considerations beyond how we way their initial testimony to decide. So, all else is equal, beyond the testimony. If we go searching for evidence and find that a little evidence for each side, such that the evidence we find is of equal strength for both sides' stories, then things have remained such that all else is equal. But, if we find strong evidence for one story and no evidence for the other, then all else is no longer equal.

since it involves accepting the conclusions of these arguments about what is rational, but denying that is what should be believed. That is certainly a conceptual possibility, but it has no bearing on whether believing the victim is rational.[90]

So, those who wish to maintain that believing the complainant and doubting the accused is irrational are going to have to pick premises. I see little reason to doubt BV1. DD1 gains support from the very same reasons that support BV1. If false reports are rare, then true reports are overwhelmingly the norm.

So, the only option to those who wish to deny the rationality of believing the victim is to deny that the statistical information in the absence of any other considerations can settle the matter of whether to believe the complainant or accused.[91] I will return to this issue in section 5. For now, I want to proceed with the positive case for thinking that believing the victim is rational and can improve the rationality of investigations.

1.2 – The Rationality of Evidence Gathering

[90] Note: the word "rational" is ambiguous. While it is common to use "rational" without any modifier, philosophers may wish to ask whether we have in mind epistemic, practical, ecological, subjective, or objective rationality. The use of "rationality" in this essay is perhaps best thought of as epistemic rationality. But debate about just what epistemic rationality is, whether it can be delineated from the practical and the ecological, etc. make it so that not much clarity is gained by using "epistemic" as a modifier. There are also ways of understanding both subjective and objective rationality, which result in understandings of "rationality" that are consonant with how the term is used here, as well as ways of understanding them both that do not fit. I cannot hope to solve the rationality wars in this essay and have left "rationality" bare throughout, convinced that charitable readers who are sensitive to these issues will be able to find a handful of things that are called "rationality" that can fit easily enough, while those who are not sensitive to what is at issue how to use the term "rationality" are unlikely to get much value from getting into the weeds with the question.

[91] It is worth pointing out that details given in testimony may affect the rationality of believing either one. But we are interested in the cases where all else is equal that would result in a he-said-she-said standoff kind of situation, where there is no other evidence than the testimony offered by the complainant and the accused. The question at issue is not whether to believe the complainants at all times, but rather whether to believe complainants over the accused when all else is equal—when we've got two competing stories with no obvious defects, no other evidence (or if we do have other evidence, what we have is of equal strength on either side), etc. Hence, the *ceteris paribus* clauses.

I turn now to the zetetic norms for investigating sexual assault.[92] To rationally gather evidence and utilize it in our thinking, three conditions must be met: *First*, you have got to get your hands on the evidence that is readily available. An investigation that does a poor job of looking for evidence is a poor investigation, and its conclusions will be poorly supported. *Second*, you have got to preserve the integrity of that evidence when you are getting your hands on it. If detectives handle their evidence poorly, they will compromise its integrity and thereby its usefulness qua evidence. *Third*, it seems like you have to critically evaluate the acquired evidence, in order to get clear on what it does and does not show.

Believing the victim policies can improve the rationality of investigations with regards to the first of these points. Victims are often in a fragile state when reporting their assaults. Victims often also face censure, negativity, scrutiny, or victim blaming when they disclose their experiences (Ahrens 2006). And so victims may have already experienced such things or be aware of the commonness of such reactions. Too much questioning or too critical questioning can cause victims to shut down. Anticipation of having to dredge up and comb through memories of their painful experiences with a fine-tooth comb can cause victims to not report. People often prefer to move past painful experiences, rather than carefully scrutinize their horrific details, thereby extracting the maximum amount of pain to be had from them. Thus, overcritical attitudes toward victims hinder the collection of available evidence. By treating the complainant as someone who has been through a horrible ordeal, and by recognizing that giving the report and facing scrutiny about details of the report can induce profound psychic pain in them, and that overcritical and uncharitable questioning can

[92] Zetetic norms are norms of inquiry. See Friedman 2020 for an overview and for an argument that zetetic norms are the only norms that matter in epistemology. See Haziza 2022 and Steglich-Petersen 2021 for some resources for reconciling traditional epistemology with zetetic epistemology.

exacerbate these problems, those investigating can do a better job of recovering the evidence stored in victims' memories. Believing victims' testimony can be one way of helping interviewers position themselves to extract the maximum amount of memory evidence, by helping to avoid causing victims to shut down and by removing disincentives for reporting. Thus, by facilitating the greater procurement of evidence, believing the victim policies stand to improve the rationality of investigations. I do not mean to claim they are the only way of effecting such positive changes. But they do seem to be a way of doing so.

Believe the victim policies have historically had a more mixed pedigree on the second point—on preserving the integrity of evidence. At least one policy implementation has gone quite poorly in this respect. Some interviewing techniques have run the risk of compromising the evidence housed in victims' memories, by running the risk of distorting or implanting memories. Such techniques have tried to make the interviewer an ally and an active component in constructing a narrative of what happened, by suggesting ways of filling in gaps or ironing out apparent inconsistencies. Doing so may help the victim make good sense of what they do remember, but it makes the interviewer's imagination a possible distorting influence in the process. This may compromise the process of reconstructing memories and may even implant false memories—memories which may then later be contradicted by witness testimony, video footage, etc.—harming the victim, undermining her reliability as a witness, and potentially even making her doubt her own reliance on her own memories. Such interviewing techniques were at the heart of the memory wars in the 90s, and some initiatives aimed at victim advocacy have frightening echoes in the

interviewing techniques that led to innocent people being falsely imprisoned as a result of implanted memories.[93]

On the other hand, an overly critical approach to interviewing the victim may also cause the victim to panic, draw hasty inferences, get confused, and do a poor job of preserving the integrity of the information housed in their own memories. A softer approach can facilitate the careful extraction that needs to be done to ensure that one is clear about every scrap of detail that is remembered, where there are gaps in what is remembered, and where there are fragments whose place in the overall story is unclear to even the person remembering.

So it seems that at least *some* policy move toward what advocates of believing the victim policies are after stands to improve the rationality of investigation. But not every policy attempting to put the *believe the victim* mantra into action is likely to improve the rationality of the investigation—as is demonstrated by one actual attempt at implementation that has been shown to be potentially very problematic. But a bad solution to a problem is no indication that the problem does not need to be solved, nor is it an indication that no good solution is available.

It is on the third point about the rationality of evidence gathering that there seems to be the biggest concern—with assessing the quality of the evidence. Even if we grant that believing the victim policies have a valuable role to play in facilitating greater evidence gathering, and perhaps even in preserving the integrity of the evidence, they may hinder investigators' ability to evaluate the quality of the evidence produced qua evidence. Arizona politician Debbie Moak's worry that such policies introduce a confirmation bias comes in here. I will tackle that concern in the next section.

[93] See Davis and Loftus 2019 for exploration.

Section 2 - The Start of Inquiry

It is the nature of grass roots movements that no one person or institution can speak for everyone involved in the movement. Few people who are worried about the testimony of victims not being given appropriate credence are likely to say that we should believe all reports of victimization at all times, no matter the complainant's credentials, no matter their story. It is fine not to believe reports of alien abduction. It is fine not to believe reports of demonic possession. It is fine not to believe sexual assault reports, when the complainant says they were assaulted by a ghost,[94] etc. But any reasonable proposed corrective to the moral failure of not giving appropriate credence to victims' testimony is going to have to say something about when to believe complainants, in order to amount to a positive suggestion about how to correct behaviors that result in the moral failing of not giving sufficient credence to victims' testimony.

The *Start by Believing* campaign,[95] created by *End Violence Against Women International*, proposes that it is at the start of inquiry, at least, that changes need to occur. As I understand the core part of their ideology, it is that believing the complainant should be the default position at the start of inquiry. The Start by Believing campaign recommends that when investigators hear a report of sexual assault, they respond by saying "I believe you."

[94] Though it may be worth taking seriously the idea that they may have been assaulted by someone who is not a ghost/alien/demon, who they mistook for a ghost/alien/demon. A mentally ill person may be truly reporting an assault, but their less than firm grip on reality may result in them making some pretty fundamental mistakes about the nature of their experience. And I certainly do not wish to suggest that we should not take seriously the testimony of mental ill people, as they are particularly vulnerable. I only mean to say that we need not throw caution to the wind and believe in ghosts so as to never question the testimony of a person reporting an assault.

[95] Startbybelieving.org

No doubt, this is validating in all the right ways.[96] I suspect most people, even just imagining themselves being in the horrible position of having to report such a thing, would say this is exactly what they would want to hear after reporting what happened to them.

But is believing at the start of inquiry rational? This section aims to defend the case that it is. In section 2.1, I consider the charge made by Debbie Moak, of the Arizona governor's office, that a policy of staring by believing would introduce a confirmation bias that would undermine the rationality and integrity of the investigation. In 2.2, I turn to a different worry—namely, that starting by believing amounts to believing ahead of the evidence.

2.1 – Confirmation Bias

Debbie Moak, of the Arizona Governor's Office of Youth, Faith and Family, issued a letter of guidance recommending that government agencies not adopt a Start by Believing approach.[97] Among Moak's concerns were the fact that investigators may not be well-positioned to defend the methodology of their investigation on the witness stand, thereby making the policy ineffective at meeting the needs of criminal prosecution. If on the witness stand people cannot effectively explain or defend how they conducted their investigation, then the case of the prosecution is liable to be undermined.

More pertinent to our concerns here is Moak's worry that starting by believing introduces a confirmation bias, which undermines the rationality and integrity of the investigation.

---

[96] There is a question here about the voluntariness of belief. Namely, given that belief is involuntary, and whether the person hearing the report believes or not is a matter beyond their control, what are people to do that either find themselves disbelieving or having no beliefs about the matter at all to do? Should they lie, and say "I believe you" all the same? I aim to address these questions in my paper on the coherency challenge.

[97] Moak 2016

My aim here is not to say anything about whether believing the complainant at the start of inquiry is going to be defensible in court or in line with U.S. law or police policy. My aim is instead to make the case that its rationality can be defended. But I make no assumptions here about connections between rationality and defensibility in court or compatibility with the laws of any government.

Confirmation bias is "seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand" (Nickerson 1998). That confirmation bias does occur in the world should be uncontroversial. That confirmation bias can at least sometimes be rationally undermining should also be fairly uncontroversial. The fact that confirmation bias is not always sufficient to rationally undermine an investigation should also be uncontroversial. If I have a confirmation bias to believe that the murder weapon is located in a suspect's car, when I have the actual murder weapon in hand, it is of little consequence that I was biased in my lead up to that moment. The evidence I have is strong enough to make the fact that I was biased in my investigation of no real consequence. So, the question we have to face here is what is sufficient to cause a confirmation bias that is rationally undermining and whether we have got a case to that effect in believing the testimony of complainants in sexual assault reports as the default position for the start of inquiry.

One may wish to say that whenever there is a belief, there is a confirmation bias. But if this is true, it cannot be the case that confirmation bias is always rationally undermining, because we always have beliefs. We form beliefs constantly. So, unless one wants to say that inquiry is always rationally undermined, because we cannot help but form beliefs antecedent to or in the course of investigation, then one will have to either say that belief is

not sufficient for bias or grant that this bias is seldom undermining. On the latter proposal, starting by believing is a case of confirmation bias, but we don't have a case that it is a rationally undermining case of confirmation bias just in virtue of it resulting in a confirmation bias. On the former proposal, such a bias is undermining, but we don't have a case that having a belief suffices to introduce a confirmation bias.

So, either way we carve up the space, we don't have a case that starting by believing is sufficient to rationally undermine investigations. That is not to say that no case could be made. It is just to say that the existence of a belief in the complainant's story at the start of inquiry is not sufficient to get the conclusion that the rationality of the investigation has been compromised.

A defender of the claim that starting by believing introduces a confirmation bias may wish to either adjust their claim or supply additional premises to earn their desired conclusion. They might adjust their claim by suggesting that starting by believing is not sufficient to introduce a bias, but that it makes bias more likely.[98] They might instead supplement their case by saying that starting by believing requires believing no matter what comes and that surely is sufficient to produce a bias. I find the latter of these suggestions wildly implausible. No one is saying we should believe no matter what.[99] The former has some plausibility, but it has yet to explain why biases would be more likely or why we could not prevent or counteract the biases without giving up on starting by believing. So it has yet to make a solid case against starting by believing.

---

[98] This actually seems to be closer to Moak's position.

[99] If one is worried here that slogans like "believe the victim" do not make that point clear, my response is that such an expectation is expecting too much from a slogan. Slogans are not profitably thought of as universal laws admitting of no exceptions. Rather, sloganeering is an activity that is generally inhospitable to hair-splitting and careful legalistic thinking. If one rejects the injunction to "believe the victim" on these grounds, one has a problem with sloganeering in general, not the content of this slogan in particular.

Whatever the story may be, I suspect that one crucial bit has been overlooked in all of this. It is that belief formation is typically automatic and liable to happen no matter what. Those investigating sexual assault reports are liable to hear the victim's report and believe, disbelieve, or have no beliefs at all, largely independently of their will. So, if beliefs are sufficient to cause or make probable bias, then the investigations are already biased. If beliefs are not sufficient to cause or make more probable bias in the investigations already, then there is not much of a case to be made that starting by believing will bias investigations.

One may push back and say that when belief formation happens automatically, it happens in response to evidence. So, it is not as rationally problematic as starting by believing the complainant. But, as I pointed out in the previous section, investigators do have evidence at the start of inquiry—namely, the complainant's testimony and that fact that false rape reports are rare. So, there is not much of an asymmetry to be had here.

Alternatively, opponents of believing the victim may wish to push back by conceding that automatic belief formation already produces biases, but claim that a start by believing position would push the biases all in one direction—in favor of the complainant. In a normal investigation, this is not so. Biases arise organically as beliefs arise organically, and they are not directed to one predetermined conclusion. So a critic may wish to say that these organic biases even out or cancel out. There is an even distribution of biases working in favor of both complainant and accused. So, they might say, it is not rationally undermining for investigations that investigators automatically form beliefs and biases resulting from them, since the biases even out. But a start by believing policy would remove this egalitarian

distribution of biases, upsetting the balance in favor of complainants and thereby upsetting the rationality of investigations.

I have a number of bones to pick with this suggestion. First, I see no reason to think biases that randomly are in favor of one party rather than another would be any less undermining, if the biases are in fact undermining. An organic distribution of bias is not the same thing as neutrality and is unlikely to have the epistemic effects of principled neutrality.

Second, if false reports are rare, then an even distribution of biases is not truth-directed. An even distribution of biases, in the absence of any other evidence, would lead to beliefs in complainant and accused accounts at about the same rate. But complainant do not lie at the same rate as those accused. So, this would lead to false beliefs 40-60% of the time.[100] Perhaps an even distribution of biases would be rational in cases where the base rate odds of either outcome being right are equal, but in cases where the odds are clearly in favor of one outcome being true, an even bias distribution seems to bias us away from the truth. If that point is not clear, just imagine an even distribution of biases supporting those who do and do not believe that Sarah will win a lottery, with a million entries and one winner to be chosen at random. Forces pulling equally in either direction, when the odds so clearly support Sarah's not winning the lottery, are not truth-directed or rational. Surely, a bias to think that one will not win the lottery is more truth-directed.

Lastly, it seems to be just plain false that the distribution of biases are equal in the typical case of the he-said-she-said standoff, where the complainant alleges and the accused

[100] In case this point is not obvious, consider that in a sample of 100 cases, at most 10 percent will be cases of false reports. That means, not believing the complainant would occur 50 times, in an egalitarian distribution of biases. Assuming that they caught the 10 lying complainants, that would make for 40 out of fifty cases where they doubted the complainant getting things wrong. If they missed all ten lying complainants, then they would get those 10, plus 50 more wrong. If the actual amount of false rape reports is lower, the final amount of cases where they get things wrong should fall within this range as well.

denies the guilt of the accused. Perhaps if we consider only confirmation biases, this is so (though I doubt it). But we also have to consider the fact that there are widespread myths about false reports being common. A survey of the attitudes of police officers found that their opinions of the occurrence of false rape reports ranged from 5% to 90% (McMillan 2018). There are common folk psychological myths about how rape victims will act, which seems to be disconnected from reality.[101] There are common stereotypes, sometimes called "rape scripts" about what rapes look like, leading less stereotypical cases to be discounted as not rape (Masters et al 2013). There are widespread myths about women wanting sexual contact even when they say they do not. There is, in fact, a culture of misogyny that favors men over women (even when the men are *known* to be guilty),[102] and police officers are overwhelmingly male.[103] Add to this the fact that those working with sexual assault victims often experience second-hand trauma, which police departments often do not take seriously, and it seems likely that many seasoned police officers are apt to be resistant to believing a rape has occurred, due to unprocessed trauma from working with past victims of rape (Crivatu, Horvath, and Massey 2023).

Perhaps a case is to be made that there exists an even distribution of biases in some police departments here or there, but it seems hard to believe such a distribution is the norm. So, we have good reason to think that the distribution of biases among those investigating sexual assaults already works in favor of the accused most often, and thereby are directed in

[101] Marie's case, discussed above, is one where people's beliefs about what a rape victim will act like are disconnected from reality. The picture we get of how someone experiencing trauma will act from movies and TV is often disconnected from the reality.

[102] For a disturbing laundry list of known rapists who went unpunished, police activity that favors the accused, and other staggering issues about rapists having an easier time than their victims, see Loofbourow 2019. For further exploration of the biases working against victims, see Crewe and Ichikawa 2021.

[103] The FBI placed the number at 87.4% male officers in 2018.

favor of what is, all else being equal, the less likely conclusion. So, even if believing the victim were to introduce a confirmation bias or direct existing confirmation biases uniformly toward the victim, it seems that doing so would actually increase the truth-directedness of investigations.

In sum, it seems as though the case for start by believing policies introducing a rationally undermining bias is not very good to start with. If belief were sufficient for a rationally undermining bias, then our investigations would already be undermined. Investigators are apt to have beliefs, whether we like it or not. The case that existing biases are non-existent or that are even in their distribution is also poor. Biases are already widely directed in favor of the accused, often even when their own guilt is admitted and widely known. The case that a start by believing policy would make our investigations biased or more even-handed is even poorer. Even if a new bias were introduced or existing biases were directed toward favoring the complainant, doing so would be a bias that favors the more likely outcome, counteracting biases that already favor the less likely outcome. So, there just is no case to be made that making believing victims the starting point of inquiry undermines the rationality of investigations. There is, on the contrary, good reason to think that starting by believing will make investigations more likely to get to the truth.

2.2 – Believing in Advance of the Evidence

A second worry about the rationality of starting by believing is that doing so requires believing in advance of the evidence. Believing in advance of the evidence is not rational, many claim[104] (whether it is or not will not be explored here).

[104] For the classic debate, see Clifford 1877 and James 1896. For notable defenses of evidentialism, see Feldman and Connee 1985 and Arpaly 2023. For defense of the claim that belief in the absence of evidence is

My response to this is that it is just plain false that believing the victim requires believing in advance of the evidence. When the investigator has heard the testimony of the complainant, they have evidence—namely, the complainant's testimony. Adding to this, they have (or at least ought to, if they are in the business of investigating sexual assault reports) the information about the rarity of false reports. In the absence of any defeaters,[105] they seem rationally entitled to belief in the veracity of victims' testimonies. They have evidence and they have no defeaters for that evidence.

One might wish to say that it is believing in advance of the total body of evidence that will come. But that is not irrational. One need not have all of the evidence available or all of the evidence that will come to have a rational belief.[106]

One might instead wish to claim that starting by believing requires dismissing defeaters in an irrational way. If the victim says something wildly implausible, flatly contradictory, or completely disconnected from reality, the investigator still has to believe! Or so they might say. But I find it implausible to think that anyone advocating a policy of making believing the complainant the starting point of inquiry actually thinks that starting by believing requires believing at the start of inquiry, no matter what. A more charitable construction of

---

rational and possible, see Rinard 2019 (and also Rinard 2015 and Rinard 2018). For defense of a stronger claim that believing *against* evidence is rational and possible, see McCormick 2014. For defense of the claim that knowing without evidence is possible, see Moon 2012.

[105] This section and the sections that follow will use "defeater" pretty liberally. I take it, with McGrath, that "Pollock's framework for defeat is now a part of the received wisdom in analytic epistemology" (McGrath 2021, p. 201) and needs no explanation. The only commitment my arguments require herein is that Platinga's (2000) controversial thesis that unjustified beliefs can be defeaters is, as many have argued (e.g. Alston 2002), false. Were that thesis true, then rape myths, which everyone should by now know to be false (see Mason 2021 on the *should know* component), would make believing the complainant irrational, so long as officers subscribed to such myths. (For classic works on the taxonomy of defeaters, see Pollock 1986 and Bergmann 1997.)

[106] An issue that I cannot get into here is the fact that belief comes in degrees. While philosophers have lately explored the idea of outright belief (e.g. Wedgwood 2012), there is serious doubt that our natural language use of the word belief actually tracks anything like outright belief—there is, on the contrary, substantial evidence that "belief is weak," meaning that people are willing to say they believe on much weaker grounds than the outright belief model predicts (see Hawthorne, Rothschild, and Spectre 2016 for exploration).

the position is that starting by believing is the default position in normal circumstances, where the person hearing the testimony has no defeaters. If someone tells you that *you* assaulted them, when you know you did not, then you are not obligated to believe them. If someone tells you that Jim assaulted them, when you know Jim was on the other side of the planet with you at the time of the assault, you are not obligated to believe the report. But if someone tells you that they were assaulted and you do not have antecedent knowledge that acts as a defeater, then it is rational to believe them.

In the following section, I'll explore whether fulfilling our duties to victims in other ways than merely just believing at the start of inquiry, might warrant further inspection of the role of defeaters and the defeasibility of believing the complainants in sexual assault reports.

3 – Believing after the Start of Inquiry

In section 1, I made an initial positive case for the claim that believing the victim policies can improve the rationality of investigations into sexual assault reports. In section 2, I examined the case for a specific policy that we might adopt—namely, to make believing the complainant the default position at the start of inquiry, all else being equal. I addressed criticisms of this policy and argued that that such a policy would not rationally undermine the integrity of an investigation. I turn now to the question of how to think about what comes after the start of inquiry. Is it the case that we have good reason to think that advocates of believe the victim policies think some further obligation to victims is still necessary and if so, is what is being advocated rational?

I repeat my cautionary statements of the previous section here. I cannot pretend to speak for the movement as a whole or to speak to the attitudes or interests of anyone and everyone who may repeat the slogan "believe the victim." But I can entertain some plausible suggestions and defend their rationality.

In order to see why we might think that there is some further obligation beyond just making believing the victim the starting point of inquiry, it may be helpful to deal with a case which has received attention in popular media. Marie was raped at knifepoint by Marc O'Leary.[107] O'Leary was a fastidious army veteran. When O'Leary took the laces from Marie's shoes, in order to bind her hands and feet, he put the shoes back in a neat and orderly fashion. When Marie's rape was investigated, the investigating detective took the orderliness of those shoes to be telling. Who would care to place the shoes back neatly, when they are about to rape someone? Why would they care about the person's shoes when they clearly don't care about the person? Why would they take the time to do this, rather than getting to the business they came here for?

The "counter-evidence" against the veracity of Marie's claims stacked up in the days to come. Marie had given slightly different stories to different people about what body parts were involved in making phone calls when she was tied up. Marie had cracked a joke the day after the rape. Marie also wanted the same kind of bedsheets that had been taken from her home by the detectives, who took them to check for DNA evidence. Marie's foster mothers thought that wanting the same bedsheet type she was raped in and enjoying a moment of levity the day after the rape were both too incongruent with what they would

[107] Marie is her middle name, but her first and last name are omitted from media discussing her case. For a detailed account of Marie's case, see Miller and Armstrong 2015.

expect from a rape victim. They phoned the detectives to report their suspicion that Marie was lying.

The detectives scheduled an interview with Marie, with the aim of getting her to admit she was lying. Marie resisted for a time, but finally said that she was willing to say she only dreamed that she was raped. The detectives pressed harder, and she finally relented and signed a confession indicating that she made it all up. A few days later, Marie came back to recant the confession, but she was threatened with jail time, and she caved. Marie was then charged for making a false report. She had to pay a fine and had to enter mandatory therapy to deal with her "lying problem." In order to keep her housing, she had to publicly tell her neighbors she had lied in a meeting. Her best friend started a website dedicated to shaming her for being a liar. Newspapers published articles shaming her for being a liar. Friends and loved ones soured on her. In short, Marie was humiliated, both publicly and privately, and left to deal with the trauma of a violent rape that she was not allowed to admit happened to her therapist, neighbors, or loved ones—all of whom now had a rather negative outlook on her character.

Marie's story is horrifying. She was only vindicated because her rapist went on to commit a series of further rapes and was eventually caught. But it is worth noting here that people did start out by believing Marie. It seems like Marie's foster mothers and friends initially believed. It seems not unlikely that the investigating detectives believed at first as well. So, if there is some belief-related defect in how Marie's tale unfolded, it does not seem as though it is explained by a failure to start by believing.

One way of responding to Marie's story might be to say that there were no belief-related defects. Perhaps, the police charging Marie and the series of public shaming rituals she was

subjected to were wrong, but the failure to believe in this instance was not rational or moral failing. It is the nature of rationality that we are not always justified in believing what is true and that we are sometimes justified in believing what is false. And perhaps this is a case of that happening. No doubt, failing to believe victims who really are victims causes a lot of harm, and Marie's case effectively demonstrates that, but perhaps that is all it demonstrates. Perhaps when it comes to belief formation and revision, regarding the veracity of Marie's story, the detectives were perfectly rational. They started by believing, but then, when they had defeaters, they stopped believing. What could be more rational than that?

While I am not unsympathetic to this line of response, and I see no a priori reason to think such horrible things could not happen as the result of faultless epistemic conduct, I do think that there were belief-related defects here. For one, the foster mothers are testament to how typical folk psychological intuitions and "rape scripts" can lead us badly astray when dealing with atypical situations. The foster mothers thought they knew how a rape victim would act, how Marie would act in particular, but they were just way off the mark. Whether they were culpable or not is debatable, but it certainly seems like an epistemic failing in the thin sense of *failure*. Whether or not Marie's foster parents were culpable, we have good reason to think that those responding to reports of sexual assault should not be prone to the kind of thinking that led them to discount Marie's testimony.

A second piece of bad folk psychology comes from the detective. The detective thought there was something off about the shoes. And there was. It was a clue to the nature of the perpetrator (he was fastidious in the way that army vets often are), but it was taken to be evidence that Marie was lying—that she had placed the shoes back herself with care, because they were her shoes. Had the detective been more willing to give Marie the benefit

of the doubt, this clue could have aided the search for the perpetrator. It was instead used to aid in talking Marie into withdrawing her report. The detective's assumptions about what a rapist would be like also led him astray. This again, seems like a failing, at least in the thin sense.

It seems to me that Marie's foster mothers and the detectives investigating her report overestimated the strength of their defeaters to her testimony. Whether they did so in a way that is culpable need not concern us here. But the evidence of Marie having been bound, of bruising to her genitals, and her testimony about the assault are all much stronger evidence for the veracity of her story than the inconsequential details that were used to attack her story. It seems to me that those who doubted Marie's testimony threw away strong evidence on the basis of weak defeaters. Everything that was used to discount Marie's story was, in fact, perfectly consistent with her story—as is evidenced by the fact that her allegations were *all true*.

And so it seems to me that there is a thin failing (at least a thin one) that extends beyond just the start of inquiry. It is not enough to just start by believing those who report an assault, you have got to be careful in what you take as a defeater of this belief. You have got to be careful in how you think about the defeaters. Take for example the different stories Marie gave about what body parts were used to make phone calls. When talking to her ex-boyfriend/best friend, she said she had to use her toes to call. This was not mentioned in stories given to detectives. But surely, when talking to someone she might have feelings for, it is not unreasonable to think she might have exaggerated a minor detail for sympathy or effect. How a teenage girl frames things when talking to a boy she is trying to impress may not be the best indicator of how she thinks things actually are. Alternatively, it could be that

she did use her toes at some point and either didn't mention this to detectives, because she forgot, or because it did not seem pressing at the time of the interview. In short, there are charitable ways of explaining this piece of "counter-evidence" that would not require fabrication of a rape. What is more, the explanations offered here seem like better explanations than the explanation that the rape was fabricated. And similar stories can be told for every other supposed piece of "evidence" counting against Marie.

And so, I think there is a good case to be made that in Marie's case, every issue that led the detectives astray resulted from a failing (at least a thin one)—not a failing of not getting the right conclusion, but of not evaluating their evidence properly. But I also think they are all-too-human failings. People are on guard against liars, as a matter of evolutionary programming. What is more, when they suspect lying, they can get even more defensive and even more suspicious, leading to hastily drawn inferences, motivated reasoning as a result of feelings of hurt at being manipulated, a desire to catch the liar in their lies, etc. It seems to me that if we are looking for a confirmation bias that rationally undermines in investigations, we need look no further than how Marie was treated. Once the detectives got it in their heads that Marie might be lying, they spun ever piece of evidence that they did not know how to make fit into counterevidence, meant to show she was lying. But if they had more carefully evaluated all of their supposed counterevidence, I think they would have found that there were plenty of good explanations on offer for them and that those explanations were all more plausible than the explanation that they were going with—that she fabricated the rape.

So, what is the takeaway from all of this? It is that investigations of sexual assault reports do require of investigators more than just starting by believing. They also require

careful and cautious evaluation of defeaters for complainants' testimony. They require considering competing explanations for the evidence and never losing sight of the fact that false rape reports are rare in the United States today. And because they are so rare, there is a greater evidential burden for discounting the testimony of a complainant offering a sexual assault report. So an explanation of a fabricated rape should not be considered as a good explanation without very good evidence of fabrication. If the point is not obvious, consider an analogy. I lose my keys frequently. It is always a possibility when I cannot find my keys that someone stole them. But until I have a lot of evidence, I should be hesitant toward jumping to the conclusion that someone broke into my apartment and stole only my keys. It is possible, yes, but unlikely. The evidential burden for believing my keys were lost by me is much less than for the unlikely scenario that someone broke into my apartment. So too, the evidential burden for thinking someone has fabricated a rape is greater, given that false reports are so unlikely, and victims seldom have incentive for fabrication.

In sum, the complainant's testimony of a sexual assault is greater evidence than it is often treated to be. Failing to recognize this point often leads to discounting of the testimony on weak grounds. There is a failure to give an appropriate level of credence to victim's testimony, but it is surely an epistemic one. Correcting for this failure is then also a matter of making investigations more rational, not less.

Section 4 – Moral Encroachment

I have defended the case that believing the victim polices are rational, and that the case can be made on strictly evidentialist grounds. In order to correct the moral failing of not giving appropriate credence to victims' reports, we need not do anything contrary to

rationality. On the contrary, failing to give sufficient credence to complainants is itself irrational. I suspect that at least some readers were expecting me to have made my case by appeal to moral encroachment—the thesis that standards of epistemic justification vary according to moral stakes. Those who find themselves drawn to the he-said-she-said standoff view, but nonetheless want to defend the rationality of believing the victim, may find themselves drawn to an appeal of moral encroachment.[108] Rima Basu (2019b), for example, has suggested that believing the victim may be a case that calls for encroachment to lower[109] the standard of justification needed to believe the victim.

My approach here has not made any appeal to moral encroachment. This is for two reasons. The first is that the thesis is controversial. Insofar as the case for believing the victim can be made without appeal to encroachment, the case for believing the victim that we have is more dialectically effective. The second reason is that I believe that, even if moral encroachment is real and we could convince everyone of its reality, the case that moral encroachment would favor believing the victim seems to me to be not so clear cut—especially as Basu frames it. I won't say it is obviously wrong. It just seems to me to be a matter about which intelligent, well-meaning, and informed people could disagree—even if they were all on board with the thesis of moral encroachment.

---

[108] Pragmatic encroachment is a thesis popularized by the work of Jason Stanley (2005). According to Stanley, our practical interests can sometimes change what level of evidence is or is not constitutive of knowledge. The thesis can easily be thought to extend to related epistemic concepts, like justification or warrant. Moral encroachment is the thesis that moral considerations, independent of our practical interests, can encroach on the standards of knowledge and related epistemic phenomena. See Moss 2018 for one of the more notably systematic treatments of the theory.

[109] Encroachment is typically thought of as *raising* standards required for knowledge or justification. The idea that it may sometimes lower standards is appealing to many, but literature on the question is still as yet undeveloped. See Pace 2011 for an early discussion, and Cassell 2024 for a recent exploration of how some cases of profiling might lower standards.

A full exploration of the issue cannot be done here. But I can gesture at the worry. Suppose the he-said-she-said standoff view is correct. In the absence of other evidence, the total evidence is equivocal, because we have two bits of testimony which are contrary. As I've argued above, I don't think this is an accurate description of the typical allegation of sexual assault, where we do have substantive evidence to support a complaint, but let us assume we are dealing with a case of the "he-said-she-said" kind. Is this an instance in which moral encroachment must be cited to defend believing the complainant, as Basu envisions, because the evidence for belief must be buttressed with moral considerations?

It seems as though the moral stakes are as follows:

	Rape report is true; accused is guilty. 50%	Rape report is false; accused is innocent. 50%	
Believe the victim	Victim gets justice; the guilty are punished	An innocent person is severely punished; a liar took advantage of the system	?
Don't believe the victim	Victim does not get justice; the guilty go free	The innocent is not punished; the system worked	?

If the conditions under which we are deciding really are 50/50, as the he-said-she-said standoff view entails, then believing the victim is going to have about the same chance of getting justice as doing grave harm. (Of course, this is supposing that believing were sufficient to attain conviction. But let us set aside this complication, since I am only gesturing at the worry.) It seems to me at least that it is worse to send an innocent person to jail in this instance than it is to fail to punish the perpetrator. It is hard to know what the likelihood of sex offenders reoffending really is, but studies on the issue place the number between 5% and 17%.[110] So the odds are in favor of no future assaults occurring, in any

[110] Langan, Schmitt, and Durose (2003) esitmates 5.3%. Hanson and Bussiere (1998) estimated 13.4%. Harris and Hanson (2022) place the number around 17%, but also note that recidivism rates vary according to

given case, even if the accused is guilty. The odds of the accused's life being destroyed if they are innocent but convicted are pretty near 100%. On the other side, if the victim does not get justice, it is a tragedy, no doubt, but it is not clear that the gap between harm caused to a victim who does get justice and one who does not is comparable to the harm caused to a falsely accused's life being destroyed. I am at least tempted to say that the amount of harm is greater in the case of the falsely accused being convicted.

Now, my point in this is not to argue for acquittal on grounds of encroachment. My point is that it is not at all clear that if you adopt the he-said-she-said standoff view that encroachment favors believing the victim. There are significant moral stakes on both sides, and a grave risk of harm on both sides, given that the decision is being made under conditions of uncertainty. If on the other hand, one rejects the he-said-she-said standoff view in favor of my claim that it is more rational to believe the victim anyway, then there is a case to be made that the overall balance of harms would favor believing the victim—lowering the standards for believing complainants and raising them for believing the accused. But for moral encroachment to deliver this result, it has to be the case that it is, all else being equal, more rational to believe the victim, independently of the effects of moral stakes on standards of belief. So, the case of believing the victim does not demonstrate the need for moral encroachment in epistemology. If, on independent grounds, we have good reason to accept moral encroachment, working out how it intersects with this issue may prove worthwhile. But we should not accept moral encroachment to solve this issue, nor should we think that it is an illuminating application of the framework.

---

certain risk factors. Offenders who a) had male victims, b) had prior offences, and c) had victims who were "young age" were 25% likely to reoffend, whereas those who did not satisfy these criteria were likely to be closer to 5%.

In sum, the case I've made here is independent of considerations of moral encroachment. The considerations which might lead us to think that this is an issue where we need to bring in moral encroachment do not obviously help the case for believing the victim in any way. Either one accepts the he-said-she-said standoff view and the encroachment considerations do not obviously help,[111] because they do not obviously favor the believing the complainant. Or one rejects the he-said-she-said standoff view, in which case encroachment is not needed in the first place. This is not to say that moral encroachment is false or that it would not apply here in some way, were it true. It is just to say that we should believe the victim whether or not encroachment is true and we should not think that encroachment, absent the case I have made without it, is going to favor believing the victim.

Section 5 – The He-said-she-said Standoff View

In section 1, I offered some reasons for rejecting the he-said-she-said standoff view, but mentioned that I would come back to it. In the previous section, I explained why the truth of the he-said-she-said standoff would render even moral encroachment unable to make the case for the rationality of believing the victim—or at least that it is far from obvious that it could do so. My case against the he-said-she-said standoff view was that it treated the truth of the complainant and the accused's testimonies in a sexual assault investigation as being, in the absence of any other evidence, equiprobable. I also flagged that a response that said justice requires being irrational and ignoring the statistics might be plausible, but that it is dialectically irrelevant, since what we are after here is whether it is rational to believe the complainant's report, and that is a response that grants me that it is rational.

[111] To reiterate, my doubt is not that encroachment does favor belief in the victim. My doubt is that the point is or can be convincingly made obvious.

We also saw in sections 2 and 3 some concrete proposals for what believing the victim might amount to. This is an account on which the default position at the start of inquiry would be starting by believing and where those conducting the investigation count the testimony of the complainant as strong evidence and handle possible defeaters with care, recognizing how easy it is to get carried away in overestimating their significance and losing sight of just how uncommon false reports are—failing to see the extra evidential burden that is required to reject the testimony of the complainant.

All of this case really seems to rest on my assumption that the statistics matter for what you should believe—insofar as we are evaluating that belief for its rationality. But one need look no further than the moral encroachment literature to find some questioning of even that assumption. Consider the following:

> Case A: Dottie is at a Republican fundraiser. She sees a Black man in a tux. Dottie gives him her drink order. The man is not waiter. He is in fact a candidate, who is raising money.[112]

Dottie believed that the Black man was a waiter. From the perspective of statistical reasoning, it is not clear that she has made any mistake. Most Black people are not republicans. At republican fundraisers, it is not uncommon for the only non-white faces to be the help staff. Supposing that Dottie knew the statistics and made this inference statistically, and not because of some racist views about Black people's proper station in life, but truly just because of known statistics, has Dottie made an epistemic mistake?

Those who endorse moral encroachment may want to say that she has made a mistake. Because the treatment of this person is a moral issue, the moral stakes were higher and so

[112] While this example is made up for this paper, examples like this, which might be called cases of *profiling,* are floating around in the moral encroachment literature. See Basu 2019a, Moss 2018, and Begby 2018. See also Cassell 2024 for exploration of cases of *positive profiling.*

standards of justification are higher. These kinds of cases are often used to make the case for why we need moral encroachment in the first place. So this was an epistemic failing, they would say, since encroachment changed the standards of justification.

I am not so sure this response helps, unless one is willing to say that the standards are raised so high that no amount of evidence could justify Dotty. If we construct the case carefully enough, we can make the odds in favor of Dottie's accuracy pretty easily. So, unless the moral encroachment advocate is going to say there is no amount of evidence that could justify Dottie's belief, we are going to be able to get versions of the case where Dottie is justified all the same. And her doing so is likely to seem to many just as unacceptable. If on the other hand, they say they will gladly allow that encroachment makes Dottie's belief unjustified under any circumstances absurdity will follow. Take for example, Case B:

> Case B: Dottie is at a Republican fundraiser. She sees a Black man in a tux. The man asks her for her drink order. Dottie believes he is a waiter and gives him her drink order.

Certainly, in Case B, Dottie would be justified in believing she is dealing with a waiter, no matter the race. It is still possible she is wrong. It could be someone hitting on her, but doing so in a way that is easily mistaken for a waiter taking a drink order. It could be someone who just wanted to know what she was ordering, so they could make up their own mind about what to order. It could be someone posing as a waiter in order to infiltrate the fundraiser. So the possibility of being wrong is still present, but surely Dottie is justified now. But if the very high probability of being a waiter now is sufficient justification, then why not in Case A? We could rig up Case A in such a way that Case A comes with even higher probabilities of waiterhood than in Case B—such as by increasing the size of the non-white waitstaff. But I suspect no matter how we rig the case, A is going to strike most

people as unacceptable and B is going to strike people as acceptable. It strikes me as highly implausible that a difference in probability is going to explain the asymmetry in our acceptability judgments.

So, if one is going to deny that Dottie is justified, I see little reason to think an appeal to moral encroachment is going to help. I suspect it is just going to press us to construct the case more carefully, so that the statistical information weighs even more heavily in favor of the inference in Case A. The only other option for holding onto the unacceptability of A is to say the standards are so high that she could never hope to meet them based on the statistics. But that will make her unjustified in thinking a person asking for her drink order is her waiter in Case B. So moral encroachment is no help here as far as I can see.

The only strategy for denying Dottie is justified in case A that I can see, whether appealing to encroachment or not, is to deny the rationality of inferences from general statistical claims to individuals. One might try something like this: Dottie is not justified in believing the Black man is a waiter, but she is justified in a weaker claim, that it is very likely he is a waiter. I see no reason why this would help, though, as it seems as though we are almost always in a situation where, if pressed, we'd say something is very likely, rather than definitely true. If you ask me any fact I am confident on, and say "are you sure?" I am likely to hedge and say "like 99% sure." I'm happy to report that as a belief. So, if Dottie's statistics say it is 99% likely that the Black man is a waiter in Case A, I fail to see how that is different than believing—as in my own case, I detect no difference between the states of believing and being 99% sure.

So, those wishing to deny the rationality of Dottie's inference seem to only have the option of saying that she can have statistical information about the likelihood of a Black

person being a waiter, but that she cannot bring it to bear on any particular Black person. And they have to say that the reason for this is that it is some kind of epistemic error. I am not optimistic that they will find one. My own take is that we have a case of genuine normative conflict. Dottie is being rational, but she is guilty of a moral failing. We have moral obligations to sometimes act in rationally suboptimal ways, even when doing so makes us vulnerable to errors. That is my take anyway. I cannot defend that take here, due to space limitations.[113,114]

The takeaway here is that the only way to defend the he-said-she-said standoff view as the rational view is to deny that the statistics about members of groups are sufficient, in the absence of any defeating evidence, to determine what it is rational to believe of a particular member of a given group. There are people who are worried about these inferences and actively trying to show why some of them are irrational.[115] So there may be hope for the he-said-she-said standoff view, if their efforts are vindicated. But I am well convinced that there is nothing wrong with the inference epistemically. Socially, practically, and morally, some of these inferences are problematic, no doubt. Again, my take is that being moral

[113] See Driver 2001 for a defense of a similar view. See Basu and Schroeder 2018 and Basu 2020 for contrary opinions on normative conflict.

[114] Another line that a moral encroachment theorist may take here is that Dottie was in fact morally and epistemically in the clear in case A, despite common intuitions to the contrary. They might say that moral encroachment raises the standards for belief higher, but that so long as we are imagining case A with the odds high enough, we should grant that Dottie has met the rational standards required. Thus, the statistics play a significant role, but they are not the only determinant of the rationality of Dottie's belief. They might thus argue on the grounds of moral encroachment that the stakes raise the standards of justification for believing the testimony of either the victim or the accused. How plausible this line of attack will be will depend on the precise calculus for standards of justification. For example, we might think that since there is more at stake for the accused, if they are innocent, we have to have really high standards. But the fact that false reports are so rare may still be enough to counteract that issue. Or it might not. The matter is sufficiently muddy, and the relevant factors about raising of stakes are so difficult to quantify that it is hard to think that moral encroachment gives an obvious answer here. Any theory of moral encroachment that has the mathematical resources to calculate the verdict precisely would be objectionable on grounds that it likely has admitted arbitrary precision.

[115] Outside of those already cited in the encroachment camp, Elizabeth Jackson (2019) has offered a novel solution, based on a different controversial thesis: belief-credence dualism.

sometimes requires relaxing our epistemic vigilance, rendering us epistemically vulnerable to error, so as to avoid moral failings that are of greater significance than the possible epistemic errors would be. But for those who do not agree, there may be hope that a purely epistemic account of the unacceptability of these inferences can be given. If so, that account may be of some help to a he-said-she-said standoff view. I, however, remain skeptical of the prospects of doing so and am happy to count the he-said-she-said standoff view as an incorrect assessment of what the ceteris paribus evidential situation is in sexual assault reports, on statistical grounds alone.

Section 6 – Conclusion

I've argued here that believing the victim, at least under some understanding of what that means, is rational. Complainants are not believed as often as it is rational to do so. Correcting the moral failing of victims' testimonies not being given appropriate credence need not require any irrationality or appeals to controversial theses like pragmatic or moral encroachment. Even on a strictly evidentialist framework, it seems as though victims are believed less often than is rational. All else being equal, it is rational to believe the complainant's report of a sexual assault over the denial of the accused. Evidence is always defeasible, so we should not say that one should believe any and every report, as we will often have defeaters. But defeaters must be carefully assessed, and the probability of alternative explanations must be assessed under the understanding that false reports are rare and that there is a greater evidential burden required to claim that such a rare occurrence has taken place.

Bibliography

Adler, J. (2002). *Belief's Own Ethics*. Cambridge: MIT Press.

Adolphs, R. and D.J. Andersen (2018). *The Neuroscience of Emotion: A New Synthesis,* Princeton: Princeton University Press.

Ahrens, C. (2006). "Being Silenced: The Impact of Negative Social Reactions on the Disclosure of Rape," *American Journal of Community Psychology,* 38:31-34.

Alston, W. (2002). "Plantinga, Naturalism, and Defeate," in James Beilby (ed), *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument against Naturalism,* Ithaca: Cornell University Press, pp. 176-203.

Armour-Garb, B., and J. Woodbridge (2010). "The Story About Propositions," *Noûs* 46(4):635-674.

Armstrong, D.M. (1968). *A Materialist Theory of Mind,* New York: Routledge.

Arpaly, N. (2023). "Practical reasons to believe, epistemic reasons to act, and the baffled action theorist," *Philosophical Issues* 33 (1): 22-32.

Audi, R. (1994). "Dispositional beliefs and dispositions to believe," Noûs 28 (4):419-34.

Audi, R. (2020). "Doxasticism: Belief and the Information-Responsiveness of Mind," *Episteme,* 17 (4), 542-562.

Balaguer, M. (1998). "Non-uniqueness as a non-problem," *Philosophia Mathematica* 6(1):63-84.

Barrett, J.L. (1999). "Theological Correctness: Cognitive Constraints and the Study of Religion" *Method and Theory in the Study of Religion,* 11:4, 29-34.

Barrett, J. L. (2004). *Why would anyone believe in God?* Walnut Creek: Altamira Press.

Barrett, L. F. (2017). *How Emotions are Made,* Boston: Houghton Mifflin Harcourt.

Basu, R. (2018). "The wrongs of racist beliefs," *Philosophical Studies* 176 (9):2497-2515.

Basu, R. (2019a). "What We Epistemically Owe To Each Other," *Philosophical Studies,* 176(4):915-931.

Basu, R. (2019b). "Radical Moral Encroachment: The Moral Stakes of Racist Belief," *Philosophical Issues*, 29, 9–23.

Basu, R. (2020). "The Specter of Normative Conflict: Does Fairness Require Inaccuracy?" In Erin Beeghly & Alex Madva (eds.), *An Introduction to Implicit Bias: Knowledge, Justice, and the Social Mind*. New York: Routledge. pp. 191-210.

Basu, R. and M. Schroeder (2018). "Doxastic Wronging," in B. Kim and M. McGrath (eds.), *Pragmatic Encroachment in Epistemology,* New York: Routledge, pp. 181-205.

Begby, E. (2021). *Prejudice: A Study in Non-Ideal Epistemology*. Oxford: Oxford University Press.

Benacerraf, P. (1965). "What Numbers Could Not Be," *Philosophical Review*, 74, 47–73.

Bennett, J. (1990). "Why is Belief Involuntary?" *Analysis*, pp. 87-107

Ben-Ze'ev, A. (2000). *The Subtlety of Emotions*, Cambridge, MA: MIT Press.

Bergmann, M. (1997). "Internalism, Externalism, and the No-Defeater Condition*." Synthese*, 110: 399-417.

Bortolotti, L. (2010). *Delusions and Other Irrational Beliefs,* New York: Oxford University Press.

Bortolotti, L. (2011). "Double Bookkeeping in Delusions: Explaining the Gap between Saying and Doing," in *New Waves in the Philosophy of Action,* ed. J. Aguilar, A. Buckareff, and K. Frankish, pp. 237-256

Bortolotti, L. (2020). *The Epistemic Innocence of Irrational Beliefs.* Oxford: Oxford University Press.

Botterill, G. and P. Carruthers (1999). *The Philosophy of Psychology,* New York: Cambridge University Press.

Boudry, M. & J. Coyne. (2016) "Disbelief in belief: On the cognitive status of supernatural beliefs," *Philosophical Psychology*, 29:4, 601-615

Broad, C.D. (1954/1971). "Emotion and Sentiment", *Journal of Aesthetics and Art Criticism*, 13(2): 203–214; reprinted in *Broad's Critical Essays in Moral Theory*, David R. Cheney (ed.), London: Allen & Unwin, 1971. doi:10.2307/425913

Buchanan, R. (2012). "Is Belief a Propositional Attitude," *Philosopher's Imprint,* Vol 12, No. 1

Camp, E. (2007). "Thinking with maps," *Philosophical Perspectives,* 21(1):145-182.

Cappelen, H and J. Dever, (2014). *The Inessential Indexical*, Oxford: Oxford University Press.

Carnap, R. (1932/3). “Psychology in Physical Language,” *Erkenntnis*, 3, pp. 107-42.

Carnap, R. (1947). *Meaning and Necessity*, Chicago: Phoenix Books, University of Chicago Press.

Cassell, L. (2024). “Moral Encroachment and Positive Profiling,” *Erkenntnis,* 89(5):1759-1779.

Chalmers, D.J. (2011). “Verbal Disputes,” *Philosophical Review,* 120(4), 515-566.

Chignell, A. (2018). "The Ethics of Belief", *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2018/entries/ethics-belief/>.

Church, A. (1951) “The need for abstract entities in semantic analysis,” in *Contributions to the Analysis and Synthesis of Knowledge,* Proceedings of the American Academy of Arts and Sciences, No. 80, pp. 100-112.

Churchland, P.M. (1981). “Eliminative Materialism and Propositional Attitudes,” *The Journal of Philosophy,* 78(2), 67-90.

Clifford, W.K. (1877/1999). “The ethics of belief”, in T. Madigan, (ed.), *The ethics of belief and other essays*, Amherst, MA: Prometheus, 70–96.

Cosmides, L. and J. Tooby (1994). “Better than rational: evolutionary psychology and the invisible hand,” *The American Economic Review,* vol. 84, no. 2: 327-332.

Cosmides, L. and J. Tooby (2000). “Consider the Source: The Evolution of Adaptations for Decoupling and Metarepresentations,” in *Vancouver Studies in Cognitive Science, Vol 10: Metarepresentations,* ed. Dan Sperber, Oxford: Oxford University Press.

Crewe, B. and J. Ichikawa (2021). “Rape Culture and Epistemology,” in *Applied Epistemology,* J. Lackey (ed.), London: Oxford University Press.

Crivatu, I., M. Horvath, and K. Massey (2023). “The Impacts of Working With Victims of Sexual Violence: A Rapid Evidence Assessment,” *Trauma, Violence, & Abuse,* 24(1):56-71.

Currie, G. and Jureidini, J. (2001), “Delusion, Rationality, Empathy,” *Philosophy, Psychiatry and Psycholog*y 8 (2–3): 159–62.

Curry, D.S. (2021) “How beliefs are like colors,” *Synthese* 199: 7889-7918. https://doi.org/10.1007/s11229-021-03144-1

D'Arms, J. and D. Jacobson (2003). "The Significance of Recalcitrant Emotion (or, Anti-quasijudgmentalism)", *Royal Institute of Philosophy Supplement*, 52: 127–45.

Damasio, A. (2000). "Thinking about Belief: Concluding Remarks", in *Memory, Brain, and Belief,* ed. Schacter and Scarry, p. 325-334.

Danziger, S., J. Levav, and L. Avnaim-Pesso (2011). "Extraneous factors in judicial decisions," *Proceedings of the National Academy of Sciences,* 108 (17), 6889-6892.

Davidson, D. (1967). "Truth and meaning," *Synthese,* 17: 304-323. https://doi.org/10.1007/BF00485035

Davidson, D. (1991). "What is Present to the Mind," *Consciousness,* vol. 1, pp. 197-213 https://doi.org/10.2307/1522929

Davis, D. and E. Loftus, (2019). "Title IX and "Trauma-Focused" Investigations: The Good, The Bad and the Ugly," *Journal of Applied Research in Memory and Cognition,* Vol. 8, 403-410.

Dawson, L. (1998). *Comprehending Cults: The Sociology of New Religious Movements,* Oxford: Oxford University Press.

de Sousa, R. (1971). "How to Give a Piece of Your Mind," *Review of Metaphysics,* 25.

de Sousa, R. (1987). *The Rationality of Emotion*, Cambridge: MIT Press.

De Zutter, A., R. Horselenberg, and P. van Koppen (2018). "Motives for Filing a False Allegation of Rape," *Archive of Sexual Behavior,* 47(2): 457-464.

Dennett, D. (1981). "Three Kinds of Intentional Psychology," In Richard Healy (ed.), *Reduction, Time, and Reality: Studies in the Philosophy of the Natural Sciences*. reprinted in *The Intentional Stance,* Cambridge: MIT Press, pp. 43-68.

Dennett, D. (1971). "Intentional systems," *Journal of Philosophy* 68(2):87-106.

Dennett, D. (1987). *The Intentional Stance,* Cambridge: MIT Press.

Dennett, D. (2022). "Am I a Fictionalist?" in *Mental Fictionalism: Philosophical Explorations,* ed. T. Demeter, T. Parent, and A. Toon, New York: Routledge.

Deonna, J. A. and F. Teroni (2012). *The Emotions: A Philosophical Introduction*, London: Routledge. Based on *Qu'est-ce qu'une émotion?*, Paris: Vrin, 2008.

Deonna, J.A. and F. Teroni (2015). "Emotions as Attitudes", *Dialectica*, 69(3): 293–311. doi:10.1111/1746-8361.12116

Department of Justice, Office of Justice Programs, Bureau of Justice Statistics (2022) *National Crime Victimization Survey.*

Driver, J. (2001). *Uneasy Virtue*. New York: Cambridge University Press.

Evans, D. (2001) *Emotion: The Science of Sentiment*, Oxford: Oxford University Press.

Evans, D. and P. Cruse (2004). *Emotion, Evolution, and Rationality*, Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198528975.001.0001

FBI Population Group (2019). "Police Employee Data." https://ucr.fbi.gov/crime-in-the-u.s/2018/crime-in-the-u.s.-2018/tables/table-74

Feldman, R. (1988). "Epistemic obligations" *Philosophical perspectives*, 2: 235–256.

Feldman, R. (2000). "The ethics of belief", *Philosophy and Phenomenological Research*, 60: 667–695.

Feldman, R. (2002). "Epistemological duties", in P. Moser (ed.), *Oxford handbook of epistemology*, New York: Oxford, 362–384.

Feldman, R. and E. Conee (1985). "Evidentialism", *Philosophical Studies*, 48: 15–34.

Ferguson, C.E., and J.M. Malouff (2016). "Assessing police classifications of sexual assault reports: A meta-analysis of false reporting rates," *Archives of Sexual Behavior,* 45: 1185-1193.

Fricker, M. (2007). *Epistemic Injustice,* London: Oxford University Press.

Friedman, J. (2020). "The Epistemic and the Zetetic," *Philosophical Review* 129 (4):501-536.

Fodor, J.A. (1968). *Psychological Explanation,* New York: Random House.

Fodor, J. A. (1975). *The Language of Thought*, Cambridge: Harvard University Press.

Fodor, J.A. (1978). "Propositional Attitudes", *The Monist*, Vol 61, No.4: 501-523.

Fodor, J. A. (1983). *The Modularity of the Mind,* Cambridge: MIT Press.

Fodor, J.A. (1987). *Psychosemantics,* Cambridge: MIT Press.

Fodor, J.A. (1985). "Fodor's Guide to Mental Representation," *Mind,* 94 (373): 76-100. https://doi.org/10.1093/mind/XCIV.373.76

Fodor, J.A. (1994). *The Elm and the Expert,* Cambridge: MIT Press.

Fodor, J. A. (2008). *LOT2: The Language of Thought Revisited*, Oxford: Clarnedon Press.

Fumerton, R. (1995). *Metaepistemology and Skepticism,* Lanham, MD: Rowman and Littlefield.

Gazzaniga, M. (1985). *The Social Brain,* New York: Basic Books.

Gendler, T. (2008a). "Alief and belief," *Journal of Philosophy*, 105(10), 634–663.

Gendler, T. (2008b). "Alief in action (and reaction)," *Mind and Language,* 23(5), 552–585.

Gertler, B. (2011). "Self-Knowledge and the Transparency of Belief," In Anthony Hatzimoysis (ed.), *Self-Knowledge*. Oxford University Press.

Ginet, C. (2001). "Deciding to Believe," in M. Steup (ed.), *Knowledge, Truth and Duty*, Oxford: Oxford University Press, pp. 63-76.

Gigerenzer, G. (2008). *Rationality for Mortals: How People Cope with Uncertainty*, Oxford: Oxford University Press.

Godfrey-Smith, P. (1998). *Complexity and the function of mind in nature*. Cambridge: Cambridge University Press.

Goldie, P. (2000). *The Emotions: A Philosophical Exploration*, Oxford: Oxford University Press. doi:10.1093/0199253048.001.0001

Goldie, P. (2010). *The Oxford Handbook of Philosophy of Emotion*, Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199235018.001.0001

Goldman, A. I. (1970). *A Theory of Human Action*. Englewood Cliffs: Princeton University Press.

Haack, S. (1997). "The ethics of belief reconsidered", in L. Hahn (ed.), *The philosophy of Roderick M. Chisholm*, LaSalle, IL: Open Court, 129–144.

Haidt, J. (2001). "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108 (4):814-834.

Hanson, R., and M. Bussière (1998). "Predicting relapse: A meta-analysis of sexual offender recidivism studies," *Journal of Consulting and Clinical Psychology, 66 (2),* 348-362.

Harris, J. and R. Hanson (2022). "Sex Offender Recidivism: A Simple Question," *Public Safety and Emergency Preparedness Canada.* https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/sx-ffndr-rcdvsm/index-en.aspx

Harris, S., Kaplan, J., Curiel, A., Bookheimer, S., Iacoboni, M., & Cohen, M. (2009). "The neural correlates of religious and nonreligious belief." *PloS one*, 4(10), e0007272.

Haselton, M.G. and D.M. Buss (2000). "Error Management Theory: A New Perspective on Biases in Cross-Sex Mind Reading," *Journal of Personality and Social Psychology,* Vol. 78, No. 1, 81-91.

Hawthorne, J., D. Rothschild, and L. Spectre (2016). "Belief is weak," *Philosophical Studies,* 173 (5): 1393-1404.

Haziza, E. (2022). "Reconciling the Epistemic and the Zetetic," *Thought: A Journal of Philosophy,* 11 (2): 93-100.

Heenan, M., and S. Murray (2006). *Study of reported rape cases in Victoria 2000-2003,* State of Victoria, Victoria Police. http://www.police.vic.gov.au/retrievemedia.asp?Media_ID=19462

Heil, John (2021). *Relations*, Cambridge: Cambridge University Press.

Helm, B. W. (2001). *Emotional Reason: Deliberation, Motivation, and the Nature of Value*, Cambridge: Cambridge University Press. doi:10.1017/CBO9780511520044

Helm, B.W. (2009). "Emotions as Evaluative Feelings", *Emotion Review*, 1(3): 248–255. doi:10.1177/1754073909103593

Helton, G. (2020). "If You Can't Change What You Believe, You Don't Believe It", *Nous,* 54:3: 501-526.

Hieronymi, P. (2006). "Controlling attitudes", *Pacific Philosophical Quarterly*, 87: 45–74.

Hieronymi, P. (2008). "Responsibility for believing", *Synthese*, 161: 357–373.

Higgins, E. (1997). "Beyond pleasure and pain," *American Psychologist.* 52:1280-1300.

Hirstein, W. (2004). *Brain Fiction: Self-Deception and the Riddle of Confabulation,* Cambridge: MIT Press.

Hunter, D. (2022). *On Believing: Being Right in a World of Possibilities*, Oxford: Oxford University Press.

Jackson, E. (2019a). "Belief and Credence: Why the Attitude-Type Matters," *Philosophical Studies,* 176 (9):2477-2486.

Jackson, E. (2019b). "How Belief-Credence Dualism Explains Away Pragmatic Encroachment," *Philosophical Quarterly,* 69(276):511-533.

James, W. (1884). "What is an Emotion?" *Mind*, 9(2): 188–205. doi:10.1093/mind/os-IX.34.188

James, W. (1896/1956). "The Will to Believe," in *The Will to believe and other essays in popular philosophy*, New York: Dover Publications, 1–31.

Johnson, D., D. Blumstein, J. Fowler, and M. Haselton (2013). "The evolution of error: error management, cognitive constraints, and adaptive decision-making biases," *Trends in Ecology and Evolution,* vol 28, 8: 474-481.

Kahneman, D. (2011). *Thinking Fast and Slow.* New York: Farrar, Straus and Giroux.

Kanin, E. (1994). "False rape allegations," *Archives of Sexual Behavior,* 23:81-92.

Kaplan, J., S. Gimbel, and S. Harris. (2016). "Neural correlates of maintaining one's political beliefs in the face of counterevidence." *Nature: Scientific Reports*, 6:39589

Kelly, L., J. Lovett, and L. Regan (2005). *A Gap or a Chasm? Attrition in Reported Rape Cases.* Home Office Research Study 293, Home Office Research Development and Statistics Directorate.

Kenny, A. (1963). *Action, Emotion and Will*, London, New York: Routledge and Kegan Paul; Humanities Press.

King, D. and A. Zimmerman (Forthcoming). "Three Theories of Belief," in J. Jong and E. Schwitzgebel (Eds.) *Belief,* Oxford: Oxford University Press.

Klein, P. (1999). "Human Knowledge and the Infinite Regress of Reasons," ed. J.Tomberlin, *Philosophical Perspectives* 13:297-325.

Kriegel, U. (2012). "Towards a New Feeling Theory of Emotion", *European Journal of Philosophy*, 22(3): 420–442. doi:10.1111/j.1468-0378.2011.00493.x

Langan, P., E. Schmitt, and M. Durose (2003). "Recidivism of sex offenders released from prison in 1984," *Bureau of Justice Statistics NCJ 198281*. Washington, DC: U.S. Department of Justice.

Lange, C. (1885/1922). *Om sindsbevægelser: et psyko-fysiologisk Studie*. Translated as *The Emotions* (along with William James "What is an Emotion?"), A. Haupt (trans.), Baltimore: Williams & Wilkins.

Levy, N. (2011). *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*, New York: Oxford University Press. doi:10.1093/acprof:oso/9780199601387.001.0001

Levy, N. (2015) “Neither Fish nor Fowl: Implicit Attitudes as Patchy Endorsements”, *Nous*, 49: 800–23.

Levy, N. (2017). “Am I a racist? Implicit bias and the ascription of racism”, *Philosophical Quarterly,* Vol. 67, no. 268, pp.534-551

Levy, N. (2018). “Showing our seams: A reply to Eric Funkhouser”, *Philosophical Psychology,* Vol 31, No. 7, pp. 991-1006.

Levy, N. (2021). *Bad Beliefs: Why They Happen to Good People*, Oxford: Oxford University Press.

Levy, N. and E. Mandelbaum (2014). “The Powers that Bind: Doxastic Voluntarism and Epistemic Obligations,” in R. Vitz (ed.) *The Ethics of Belief,* Oxford: Oxford University Press.

Lewis, D.K. (1972). “Psychophysical and theoretical identifications,” *Australasian Journal of Philosophy* 50 (3):249-258.

Lisak, D., L. Gardinier, S.C. Nicksa, and A. M. Cote (2010). “False Allegations of Sexual Assault: An Analysis of Ten Years of Reported Cases,” *Violence against Women* 16 (12): 1318–34.

Lisdorf, A. (2007). “What’s HIDD’n in the HADD?” *Journal of Cognition and Culture*, 7:3, 341-353.

Lonsway, K., J. Archambault, and D. Lisak (2009). “False reports: Moving beyond the issue to successfully investigate and prosecute non-stranger sexual assault,” *The Voice,* 3(1):1-11.

Loofbourow, L. (2019). “Why Society Goes Easy on Rapists,” *Slate.com*. https://slate.com/news-and-politics/2019/05/sexual-assault-rape-sympathy-no-prison.html

Luhrmann, T.M. (2018). “The Faith Frame” *Journal of Contemporary Pragmatism* 15: 1-17.

Lycan, W. (1986). “Tacit Belief,” in *Belief: Form, Content, and Function,* ed. R. Bogdan, Oxford: Oxford University Press.

Lycan, W. (1988). *Judgment and Justification,* Cambridge: Cambridge University Press.

Maij, D.L.R., H.T. van Schie, and M. van Elk. (2019). “The boundary conditions of the hypersensitive agency detection device: an empirical investigation of agency detection in threatening situations.” *Religion, Brain, and Behavior,* vol. 9: No. 1, 23-51.

Mandelbaum, E. (2013). “Against Alief”, *Philosophical Studies,* 165(1): 197-211.

Mandelbaum, E. (2014). “Thinking is believing”, *Inquiry,* 57:1, 55-96.

Mandelbaum, E. (2016). “Attitude, Inference, Association: On the Propositional Structure of Implicit Bias”, *Nous*, 50: 629–58.

Mandelbaum, E. (2018). “Troubles with Bayesianism: An introduction to the psychological immune system,” *Mind and Language* 34 (2):141-157.

Mason, E. (2019). *Ways to Be Blameworthy: Rightness, Wrongness, and Responsibility,* Oxford: Oxford University Press.

Mason, E. (2021). “Rape, Recklessness, and Sexist Ideology,” In George I. Pavlakos & Veronica Rodriguez-Blanco (eds.), *Agency, Negligence and Responsibility*. New York: Cambridge University Press.

Masters, T., E. Casey, E. Wells, and D. Morrison (2013). “Sexual scripts among young heterosexually active men and women: Continuity and change,” *Journal of Sex Research,* 50(5): 409-420.

Matthews, R.J. (1994). “The Measure of Mind,” *Mind,* Vol 103, No. 410, pp. 131-146. https://doi.org/10.1093/mind/103.410.131

Matthews, R.J. (2007). *The Measure of Mind,* New York: Oxford University Press.

McCormick, M. (2014). *Believing Against the Evidence: Agency and the Ethics of Belief,* New York: Routledge.

McKay, R.T. and D.C. Dennett (2009). “The evolution of misbelief,” *Behavioral and Brain Sciences,* 32, 493-561.

McGrath, M. (2021). “Undercutting Defeat: When it Happens and Some Implications for Epistemology,” In Jessica Brown & Mona Simion (eds.), *Reasons, Justification, and Defeat*. Oxford: Oxford University Press. pp. 201-222.

McNamara, J., S. McDonald, J. Lawrence (2012). “Characteristics of false allegation adult crimes,” *Journal of Forensic Sciences,* 57:643-646.

McMillan, L. (2018). “Police officers’ perception of flase allegations of rape,” *Journal of Gender Studies,* 27: (1): 9-21.

Mercier, H. (2020). *Not Born Yesterday: The Science of Who We Trust and What We Believe.* Princeton: Princeton University Press.

Miller, G. (2003). “The Cognitive Revolution: A Historical Perspective,” *Trends in Cognitive Sciences*, 7(3).

Miller, T., and K. Armstrong. (2015). “An Unbelievable Story of Rape,” *Pro-Publica: The Marshall Project.* https://www.propublica.org/article/false-rape-accusations-an-unbelievable-story

Millikan, R. (1990). “The Myth of the Essential Indexical,” *Noûs*, 24: 723–734. https://doi.org/10.2307/2215811

Millikan, R.G. (1984). *Language, thought and other biological categories,* Cambridge: MIT Press.

Miyazono, K. (2019). *Delusions and Beliefs.* New York: Routledge.

Mlodinow, L. (2022). *Emotional: How Feelings Shape Our Thinking,* New York: Vintage.

Moak, D. (2016). “Guidance: Start by Believing,” *Guidance letter from the Arizona Governor’s Office* http://www.prosecutorintegrity.org/wp-content/uploads/2019/10/AZ-Governors-Commission-on-SBB.pdf

Moon, A. (2012). “Knowing Without Evidence,” *Mind* 121(482): 309-331.

Moon, A. (2017). “Beliefs do not come in degrees,” *Canadian Journal of Philosophy,* 47 (6): 760-778.

Moore, J. (1999). “Propositions, Numbers, and the Problem of Arbitrary Identification,” *Synthese*, 120, 229–263.

Morton, A. (2010) “Epistemic emotions”, in In Peter Goldie (ed.), *The Oxford Handbook of Philosophy of Emotion*. New York: Oxford University Press.

Moss, S. (2018). “Moral Encroachment,” *Proceedings of the Aristotelian Society,* 118(2):117-205

National Institute of Justice & Centers for Disease Control (1998). *Prevention, Prevalence, Incidence and Consequences of Violence Against Women Survey.*

Neu, J. (2000). *A Tear is an Intellectual Thing: The Meanings of Emotion*, Oxford, New York: Oxford University Press.

Nichols, S. and S. Stich (2003). *Mindreading,* Oxford: Oxford University Press.

Nickerson, R. (1988). “Confirmation Bias: A Ubiquitous Phenomenon in Many Guises,” *Review of General Psychology,* Vol 2, No. 2, 175-220

Nussbaum, M. C. (2001). *Upheavals of Thought: The Intelligence of Emotions*, Cambridge: Cambridge University Press. doi:10.1017/CBO9780511840715

Owens, D. (2000). *Reason Without Freedom: The Problem of Epistemic Normativity,* New York: Routledge.

Pace M (2011). "The epistemic value of moral considerations: Justification, moral encroachment, and James' 'will to believe,'" *Noûs* 45(2):239–268.

Paul, S. and J. Morton (2018). "Believing in Others," *Philosophical Topics* 46 (1):75-95.

Perry J. (1979). "The Problem of the Essential Indexical," *Noûs* 13 (1):3-21. https://doi.org/10.2307/2214792

Pitt, D. (2020). "Mental Representation," *The Stanford Encyclopedia of Philosophy*

Plantinga, A. (2000). *Warranted Christian Belief,* New York: Oxford University Press.

Pollock, J. (1986). *Contemporary Theories of Knowledge,* Savage: Rowmann and Littlefield

Porot, N. and E. Mandelbaum (2020). "The Science of Belief: A Progress Report," *Wires,* 1.

Poslasjko, K. (2022). "The Lycan-Stich Argument and the Plasticity of "Belief"," *Erkenntnis,* 87:1257-1273.

Powers, L.H. (1978). "Knowledge by Deduction," *Philosophical Review,* 87.

Prinz, J. (2004). *Gut Reactions: a Perceptual Theory of Emotion*, Oxford: Oxford University Press.

Quilty-Dunn, J. and E. Mandelbaum (2018). "Against dispositionalism: belief in cognitive science", *Philos Sutdies* 175: 2352-2372

Quilty-Dunn, J., N. Porot, and E. Mandelbaum (Forthcoming). "The Best Game in Town: The Re-Emergence of the Language of Thought Hypothesis Across the Cognitive Sciences," *Behavioral and Brain Sciences.*

Quine, W.V.O. (1953). "On what there is," In *From a Logical Point of View*, Cambridge, Mass.: Harvard University Press. pp. 1-19.

Quine, W.V.O. (1956). 'Quantifiers and propositional attitudes', *Journal of Philosophy*, 53: 177– 187. https://doi.org/10.2307/2022451

Quine, W.V.O. (1970). *Philosophy of Logic,* Englewood Cliffs: Prentice Hall.

Ramsey, F.P. (1931). *The Foundations of Mathematics*, London: Kegan Paul.

Rey, G. (2007). "'Meta-atheism: Religious Avowal as Self-Deception," in Louise M Antony (ed.), *Philosophers without Gods: Meditations on Atheism and the Secular Life*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780195173079.003.0019

Richardson, R. (1981). "Internal Representation: Prologue to a Theory of Intentionality," *Philosophical Topics,* 12.

Rinard, S. (2015). "Equal Treatment for Belief," *Philosophy and Phenomenological Research,* 94(1), 121-143.

Rinard, S. (2018). "Believing for Practical Reasons," *Nous* (4):763-784.

Rinard, S. (2019). "Equal treatment for belief," *Philosophical Studies,* 176(7), 1923-1950.

Roberts, R. (2003). *Emotions: An Essay in Aid of Moral Psychology*, New York: Cambridge University Press. doi:10.1017/CBO9780511610202

Rokeach, M. (1964). *The Three Christs of Ypsilanti.* New York: New York Review Books.

Rorty, A. O. (1980). *Explaining Emotions*, Los Angeles: University of California Press, 103–126.

Rosen, G. (2004). "Skepticism About Moral Responsibility", *Philosophical Perspectives*, 18: 295–313. doi:10.1111/j.1520-8583.2004.00030.x

Rowbottom, D. (2007). '"In-Between Believing" and Degrees of Belief', *Teorem*a 26, pp. 131–7.

Rousselet, M., O. Duretete, J.B. Hardouin, and M. Grall-Bronnec (2017). "Cult membership: What factors contribute to joining or leaving," *Psychiatry Research,* vol. 257: 27-33.

Russell, B. (1910). "On the nature of truth and falsehood," In *Philosophical Essays*, Longmans, Green.

Russell, J.A. and L.F. Barrett (1999). "Core affect, prototypical emotional episodes, and other things called emotion: dissection the elephant," *J Pers Soc Psychol,* May; 76(5): 805-819.

Russell, J.A. (2003). "Core affect and the psychological construction of emotion," *Psychology Review,* Jan;110(1): 145-172.

Ryle, G. (1949). *The Concept of Mind*, Chicago: University of Chicago Press.

Salmon, N. (1986). *Frege's Puzzle,* Cambridge: MIT Press.

Salmon, N. (2001). "The very possibility of language: A sermon on the consequences of missing Church," In C. Anthony Anderson & Michael Zelëny (eds.), *Logic, Meaning, and Computation: Essays in Memory of Alonzo Church.* Kluwer Academic Publishers.

Schacter, D.L., and E. Scarry. (2000). *Memory, Brain, and Belief,* Cambridge: Harvard University Press

Schiffer, S. (1981). "Truth and the Theory of Content," in H. Parret and J. Bouveresse (eds.) *Meaning and Understanding,* Berlin: de Gruyter.

Schwitzgebel, E. (2002). "A Phenomenal, Dispositional Account of Belief," *Nous,* 36 (2): 249-275

Schwitzgebel, E. (2010). "Belief", *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/ entries/belief/>

Schwitzgebel, E. (2010). "Acting Contrary to Our Professed Beliefs or the Gulf Between Occurrent Judgment and Dispositional Belief," *Pacific Philosophical Quarterly* 91 (4):531-553.

Schwitzgebel, E. (2013). "A Dispositional Approach to Attitudes: Thinking Outside the Belief Box," in *New Essays on Belief: Constitution, Content, and Structure,* ed. N. Nottelmann. New York: Palgrave Macmillan, 75-99.

Schwitzgebel, E. (2021). "The Pragmatic Metaphysics of Belief," *The Fragmented Mind,* ed. C. Borgoni, D. Kindermann, and A. Onofri. New York: Oxford University Press.

Sellars, W. (1958). Contribution to the "Chisholm-Sellars Correspondence on Intentionality" (3 August 1956), in H. Feigl, M. Scriven and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science,* Vol. 1 (University of Minn. Press); reprinted in *Science Perception and Reality,* New York: Humanities Press.

Shah, N. and J.D. Velleman (2005). "Doxastic Deliberation," *The Philosophical Review*, Vol. 114, No. 4, : 497-534.

Slone, D. J. (2004). *Theological Incorrectness: Why Religious People Believe What They Shouldn't.* Oxford: Oxford University Press.

Solomon, R. C. (1976). *The Passions*, Garden City, New York: Doubleday Anchor.

Solomon, R. C. (1980). “Emotions and Choice”, in Rorty 1980a: 251–81.

Solomon, R.C. (2003). *Thinking About Feelings: Philosophers on Emotions*, (Series in Affective Science), Oxford; New York: Oxford University Press.

Solomon, R.C. (2008). “The Philosophy of Emotions”, in Lewis, Haviland-Jones, & Barrett 2008: 3–16.

Sperber, D. (1997). “Intuitive and reflective beliefs,” *Mind and Language* 12(1): 67-83.

Sperber, D. and H. Mercier. (2017). *The Enigma of Reason*. Cambridge: Harvard University Press.

Spohn, C., C. White, and K. Tellis (2014). “Unfounding Sexual Assault: Examining the Decision to Unfound and Identifying False Reports,” *Law and Society Review,* 48(1): 161-192.

Stalnaker, R.C. (1976a). “Possible Worlds,” *Noûs* 10 (1):65-75.

Stalnaker, R.C. (1976b). “Propositions,” In Alfred F. MacKay & Daniel D. Merrill (eds.), *Issues in the Philosophy of Language: Proceedings of the 1972 Colloquium in Philosophy*. New Haven and London: Yale University Press. pp. 79-91

Stalnaker, R.C. (2008). *Our Knowledge of the Internal World,* Oxford: Oxford University Press.

Stanley, J. (2005). *Knowledge and Practical Interests,* New York: Clarendon Press.

Steglich-Petersen, A. (2021). “An instrumentalist unification of zetetic and epistemic reasons,” *Inquiry: An Interdisciplinary Journal of Philosophy.*

Stein, A. (2017). *Terror, Love, and Brainwashing,* New York: Routledge.

Stich, S. (1978). “Beliefs and Subdoxastic States”, *Philosophy of Science,* Vol 45, No.4. pp. 499-518.

Stich, S. (1983) *From Folk Psychology to Cognitive Science: The Case Against Belief,* Cambridge: MIT Press.

Stich, S. (1996). *Deconstructing the Mind,* Oxford: Oxford University Press.

Stoljar, D. (Forthcoming). “In Praise of Poise”, in *Themes from Block,* Cambridge: MIT Press

Strawson, G. (1994/2009). *Mental Reality,* Cambridge: MIT Press.

Sudduth, M. (2008). "Defeaters in Epistemology," *Internet Encyclopedia of Philosophy.*

Tappolet, C. (2000). *Emotions et Valeurs*, Paris: Presses Universitaires de France.

Tappolet, C. (2010). "Emotions, Action, and Motivation: the Case of Fear", in Goldie 2010: 325–345.

Tappolet, C. (2016). *Emotions, Values, and Agency*, Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199696512.001.0001

Tetlock, P. (1998). "Social psychology and world politics," in *Handbook of Social Psychology,* eds. Gilbert et al., New York: McGraw Hill, pp. 868-912.

van Elk, M. (2013). "Paranormal believers are more prone to illusory agency detection than skeptics." *Consciousness and Cognition,* 22(3), 1041–1046.

Van Inwagen, P. (1996). "It is wrong, everywhere, always, and for anyone, to believe anything upon insufficient evidence", in J. Jordan and D. Howard-Snyder (eds.), *Faith, freedom and rationality*, Lanham, MD: Rowman and Littlefield, 137–153.

Van Leeuwen, N. (2014). "Religious credence is not factual belief," *Cognition,* 133:698-715.

Van Leeuwen, N. (2018). "The Factual Belief Fallacy," *Contemporary Pragmatism,* eds. T. Coleman and J. Jong, pp. 319-343

Van Leeuwen, N. (2021). "Does *Think* Mean the Same Thing as *Believe?* Linguistic Insights Into Religious Cognition," *Psychology of Religion and Spirituality,* Vol 13, No.3, 287-297.

Van Leeuwen, N. (forthcoming). "The Trinity and the Light Switch: Two Faces of Belief," In Eric Schwitzgebel & Jonathan Jong (eds.), The Nature of Belief. Oxford University Press.

Velleman, J.D. (2005). *The Possibility of Practical Reason*, Oxford: Oxford University Press.

Wedgwood, R. (2012). "Outright Belief," *Dialectica,* 66(3):309-329.

Westbury, C., and D.C. Dennett. (2000). "Mining the Past to Construct the Future: Memory and Belief as Forms of Knowledge", in *Memory, Brain, and Belief,* ed. Schacter and Scarry, p. 11-32.

Williams, B. (1973). "Deciding to believe," In *Problems of the Self*, Cambridge: Cambridge University Press. pp. 136--51.

Williams, J. (1981). “Justified Belief and the Infinite Regress Argument,” *American Philosophical Quarterly,* XVIII, no 1.

Wittgenstein, L. (1958). *Philosophical Investigations*, G.E.M. Anscombe (trans.), Englewood Cliffs, NJ: Basil Blackwell.

Zamulinski, B. (2002). “A re-evaluation of Clifford and his critics”, *Southern Journal of Philosophy*, 40: 437–457.

Zimmerman, A. (2007). “The Nature of Belief”, *Journal of Consciousness Studies*, 14, N.11:61–82.

Zimmerman, A. (2018). *Belief: A Pragmatic Picture,* Oxford: Oxford University Press.

Zimmerman, M. (2022). *Ignorance and Moral Responsibility,* Oxford: Oxford University Press.