

JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY

VOLUME XXX · NUMBER 7

December 2025

ARTICLES

- 1003 Reframing Epistemic Partiality:
A Case for Acceptance
Laura K. Soter
- 1041 Elicitory Structural Power and Agential Power:
An Outline and Defense
Arash Abizadeh
- 1071 Being Wronged and Understanding Moral
Wrongness
Daniel Vanello
- 1100 A Hedonic Subjectivism
Daniel Pallies
- 1125 Freedom of Gender
Rach Cosker-Rowland
- 1171 Why Contractualism Cannot Accept
Equal Treatment for Equal Statistical Loss
Jay Zameska

JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY

VOLUME XXX · NUMBER 7

December 2025

ARTICLES

- 1003 Reframing Epistemic Partiality:
A Case for Acceptance
Laura K. Soter
- 1041 Elicitory Structural Power and Agential Power:
An Outline and Defense
Arash Abizadeh
- 1071 Being Wronged and Understanding Moral
Wrongness
Daniel Vanello
- 1100 A Hedonic Subjectivism
Daniel Pallies
- 1125 Freedom of Gender
Rach Cosker-Rowland
- 1171 Why Contractualism Cannot Accept
Equal Treatment for Equal Statistical Loss
Jay Zameska

The *Journal of Ethics and Social Philosophy* (*JESP*) is a peer-reviewed online journal in moral, social, political, and legal philosophy. The journal is founded on the principle of publisher-funded open access. There are no publication fees for authors, and public access to articles is free of charge. Articles are typically published under the CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL-NODERIVATIVES 4.0 license, though authors can request a different Creative Commons license if one is required for funding purposes. Funding for the journal has been made possible through the generous commitment of the Division of Arts and Humanities at New York University Abu Dhabi.

JESP aspires to be the leading venue for the best new work in the fields that it covers, and it is governed by a correspondingly high editorial standard. The journal welcomes submissions of articles in any of these and related fields of research. The journal is interested in work in the history of ethics that bears directly on topics of contemporary interest, but does not consider articles of purely historical interest. It is the view of the editors that the journal's high standard does not preclude publishing work that is critical in nature, provided that it is constructive, well argued, current, and of sufficiently general interest. *JESP* also endorses and abides by the Barcelona Principles for a Globally Inclusive Philosophy, which seek to address the structural inequality between native and nonnative English speakers in academic philosophy.

JESP publishes articles, discussion notes, and occasional symposia. Articles normally do not exceed 12,000 words (including notes and references). *JESP* sometimes publishes longer articles, but submissions over 12,000 words are evaluated according to a proportionally higher standard. Discussion notes, which need not engage with work that was published in *JESP*, should not exceed 3,000 words (including notes and references). *JESP* does not publish book reviews.

Papers are published in PDF format at <https://www.jesp.org>. All published papers receive a permanent DOI and are archived both internally and externally.

Editors-in-Chief

Sarah Paul
Matthew Silverstein

Associate Editors

Rima Basu	Elinor Mason
Saba Bazargan-Forward	Simon Căbulea May
Brian Berkey	Tristram McPherson
Ben Bramble	Hille Paakkunainen
James Dreier	Sam Shpall
Julia Driver	Kevin Toh
Alex Gregory	Mark van Roojen
Christie Hartley	Han van Wietmarschen
Renée Jørgensen	Kenneth Walden
Anthony Laden	Lori Watson
Coleen Macnamara	Jonathan Way

Symposia Editor

Errol Lord

Managing Editor

Chico Park

Copyeditor

Lisa Y. Gourd

Proofreader

Susan Wampler

Typesetter

Matthew Silverstein

Editorial Board

Elizabeth Anderson

David Brink

John Broome

Joshua Cohen

Jonathan Dancy

John Finnis

Leslie Green

Karen Jones

Frances Kamm

Will Kymlicka

Matthew Liao

Kasper Lippert-Rasmussen

Stephen Perry

Philip Pettit

Gerald Postema

Henry Richardson

Thomas M. Scanlon

Tamar Schapiro

David Schmidtz

Russ Shafer-Landau

Tommie Shelby

Sarah Stroud

Valerie Tiberius

Peter Vallentyne

Gary Watson

Kit Wellman

Susan Wolf

REFRAMING EPISTEMIC PARTIALITY

A CASE FOR ACCEPTANCE

Laura K. Soter

YOUR FRIEND has been accused of something serious, and the evidence is not in their favor. Or perhaps they are undertaking a difficult endeavor, and their chances of success look slim. When the epistemic chips are down, can friendship make demands on our beliefs?¹ The default response is no. Traditional evidentialism in epistemology holds that rational agents ought to believe only as their evidence warrants. In cases where being friends with someone is not evidentially relevant to a belief regarding them, the traditional view maintains that one's beliefs ought not change merely because someone is a friend. Perhaps we ought to alter our behavior regarding our friends or even practices of inquiry and evidence gathering—but not our actual beliefs. *Contra* this default view, some have argued that friendship can make demands not only on how we treat our friends but also on how we believe about them.² These “epistemic partialists” argue that in at least some cases, we can owe our friends *belief against the evidence*—claiming that friendship is manifested not only in our actions but also in our thoughts.

Each perspective seems to get something right (and something wrong). Partialism captures the psychological dimension of friendship: we really do seem to care not only how our friends act but also how they think about us—even (or perhaps especially) when things do not look good. But partialism faces challenges regarding its prescription to believe for moral reasons and worries about preserving honesty and authenticity in friendship. Traditional evidentialism avoids these issues—though (arguably) at the cost of a story about the cognitive demands of friendship.

- 1 I follow the literature in talking about friendship, though the discussed considerations plausibly apply to a variety of close relationships. Some may find the motivations of partialism even more compelling for relationships such as romantic partnerships or parent-child relationships. For these reasons, Dormandy situates her discussion of doxastic partiality in terms of love rather than friendship (“Loving Truly,” 218).
- 2 Notably, Stroud, “Epistemic Partiality in Friendship”; and Keller, “Belief for Someone Else’s Sake” and “Friendship and Belief.”

Here, I propose an account that carves a middle ground: friendship gives us reasons not of belief but of *acceptance*. I draw on a specific account of acceptance understood as regulating the characteristic role that beliefs play in guiding cognition, reasoning, and action—a profile I develop and motivate by highlighting its structural analogy to familiar mechanisms of emotion regulation. I argue that acceptance can capture the cognitive dimension of friendship while avoiding worries about positing obligations to believe against the evidence and prescribing direct epistemic irrationality.

This article proceeds as follows. First, I lay out the two poles of the epistemic partiality debate (section 1): simple evidentialism and full-blown partialism. I then introduce my account of acceptance (section 2) and propose it as a novel position within the debate (section 3). Subsequently, I more thoroughly canvass the various objections against the standard views and argue that acceptance avoids these challenges (section 4). I then address some worries (section 5) and conclude (section 6).

This article has three interrelated motivations: (1) a *normative* motivation to follow the literature in aiming to give an account of what good friends *should* do when the epistemic chips are down; (2) *diagnostic* motivation to give a psychologically plausible model of the cognitive mechanisms we may actually use in such cases; and, because within the ethics of belief, acceptance is often dismissed out of hand as an unsatisfying cousin of belief, (3) a *programmatic* motivation to demonstrate that a thorough analysis of the psychological dynamics of acceptance can vindicate its theoretical usefulness. Crucially, my proposal is a fundamentally conditional one: *if* there are indeed times when considerations of friendship conflict with epistemic rationality, then (I argue) acceptance is our best resource in these cases. My aim here is not to offer a novel defense of the antecedent.

1. EVIDENTIALISM AND PARTIALISM

Consider the following case.

Cheating: Mateo has good evidence that his best friend Shelby may have cheated on her final statistics exam. A reliable teaching assistant insists she saw Shelby staring suspiciously at her inner arm, where she now has a smudge of ink, too obscured to tell what was written. Shelby has gotten in trouble for cheating once in the past and lied about it at the time. And Shelby's performance on this final exam was quite a bit better than her performance on past exams in the class. Although Shelby told the teaching assistant she studied extra hard, Mateo has been busy with

his own finals and cannot verify that.³ Mateo feels torn: he feels his evidence suggests that Shelby likely cheated, but he also feels he owes it to her as her friend to believe she did not.

What we want from Cheating is a case of the following shape: an agent takes themselves to have evidence supporting a particular belief state, but we can imagine them feeling pulled from the perspective of friendship to believe something different. Various cases have been proposed in the literature, for instance regarding a friend's chances of success in a difficult endeavor or whether a friend spurned a romantic partner.⁴ Readers should pick whatever case they find most compelling: perhaps your friend is entering their fourth stint in rehab, and you feel you owe it to them to believe this time they will succeed (despite no notable differences in circumstance); perhaps your friend has accused someone you like of some misconduct, and you think the evidence is ambiguous, but you feel you owe it to them to believe their claim despite your doubts. Notably, the most compelling examples are those that involve not everyday unfavorable beliefs about our friends but rather more exceptional cases in which a friend finds themselves in an especially difficult situation and in need of a friend's support.

The question of interest is: What should Mateo, *qua* friend, believe about Shelby?

1.1. Two Poles of the Debate

The possible responses to this question can be positioned along a spectrum whose endpoints represent the two clearest conceptual poles of the debate. On one end is *simple evidentialism*; on the other, *full-blown partialism*. (As I discuss below, various subsequent views have been offered that fall somewhere in between these poles.)

Simple evidentialism represents the default epistemological response to our question of interest. Simple evidentialism holds that there is nothing epistemically special about cases like Cheating; what we should believe is, as always, determined by the norms of epistemic rationality. According to the traditional evidentialist (or "purist") framework, only evidential reasons (i.e., information that bears on the likely truth or falsity of a proposition) are (rational) reasons

3 I focus for simplicity on cases where the friend's testimony is not part of the evidence. Goldberg argues that friends' testimony can provide us with strong (though not insurmountable) evidence, especially in high-stakes cases, because lying would compromise trust and damage friendship ("Against Epistemic Partiality in Friendship," 2227–34). But sometimes we lack testimony, and sometimes testimony cannot settle the matter (such as in beliefs about future success). (Though Goldberg claims his explanation applies equally well to cases that do not involve testimony, his exploration of such cases is limited.)

4 Keller, "Friendship and Belief," 331–32; and Stroud, "Epistemic Partiality in Friendship," 508.

to believe.⁵ Mateo ought to believe in Cheating what he ought to believe in any other situation: what his evidence warrants—in this case, that Shelby likely cheated. Their friendship is not evidentially relevant to whether she cheated; thus, there is no difference between what Mateo ought to believe *qua* friend and *qua* epistemically rational agent (hence the ‘simple’ label: on this view, there is no special answer to the question of what to believe).⁶

Proponents of simple evidentialism hold that any norms of friendship bear only on action, not belief. We might have reason to support our friends in speech and action: to say we have their backs or do not think they cheated, to defend them to others, and so on. But such acts are *in spite of* what we believe—and like good rational evidentialists, our beliefs are still determined (only) by our evidence. Simple evidentialism says, at most, for the sake of your friend, you should act as if *p*—but do not actually believe *p* (when *p* is unsupported by your evidence).

An alternative perspective has emerged arguing that friendship can shape how we ought to *believe* about our friends. The strongest version of this view represents the other end of our spectrum: *full-blown partialism* holds that in the cases of interest, good friendship can involve *belief against the evidence*.⁷ Against the background of evidentialism, partialists propose that “there are cases in which an agent cannot meet both the highest standards of friendship and the highest standards of epistemic responsibility.”⁸ Because motivations of friendship, in these cases, have nothing to do with making our beliefs more accurate, what we ought to believe *qua* friend can be different from what we ought to believe *qua* epistemically rational agent.⁹ Regarding Cheating, a partialist might say Mateo owes it to Shelby to believe her innocence (or at least not believe

5 E.g., Feldman and Conee, “Evidentialism”; and Shah, “A New Argument for Evidentialism.”

6 A related claim here is that friendship and other practical motivations are the *wrong kinds of reasons* for belief. For discussion of this in the context of belief about others, see Enoch, “What’s Wrong with Paternalism.” For a defense of the notion, see Hieronymi, “The Wrong Kind of Reason.” And for an overview, see Gertken and Kiesewetter, “The Right and the Wrong Kind of Reasons.”

7 Keller, “Friendship and Belief”; and Stroud, “Epistemic Partiality in Friendship.” See also Baker, “Trust and Rationality”; Hazlett, *A Luxury of the Understanding*; Rioux, “On the Epistemic Costs of Friendship”; and Woodcock, “If Epistemic Partialism Is True, Don’t Tell Your Friends.”

8 Keller, “Friendship and Belief,” 330.

9 We sometimes believe more favorably about our friends than a detached observer would because we have more evidence about them; the cases of interest are those where we owe our friends epistemic and doxastic duties that are not justified by this further evidence.

her guilt), despite the unfavorable evidential situation.¹⁰ Partialists take these doxastic patterns to be deeply important to (even constitutive of) friendship: friendship is manifested through not just how we treat our friends but how we believe about them. Crucially, for a full-blown partialist, it is the *content* of the belief that matters—such that sometimes, when the evidential situation is bad, the beliefs we owe our friends are not be supported by our evidence.¹¹

I discuss below various intermediate positions that have been offered between these two poles of the debate. But first, let us discuss some of the strengths and weaknesses of these anchor positions; this will highlight what features intermediate accounts aim to capture and what problems they aim to avoid.

1.2. *Strengths and Weaknesses*

Simple evidentialism and full-blown partialism each capture and each miss something significant about the landscape of cases like Cheating. (I offer an overview of the worries here; we revisit them in detail in section 4.) Simple evidentialism highlights that the importance of epistemic rationality does not go out the window in contexts of friendship. To help us successfully navigate the world, our beliefs need to be evidence responsive; locating the domain of friendship only in external action but not in belief avoids consequences of taking on irrational beliefs. But if simple evidentialism ends up prescribing outwardly acting one way towards our friends while internally believing another, worries arise about deception and sincerity. If I ought to tell my friend that she is going to do well in her upcoming audition even if I do not believe it, or ought to defend her innocence to others while privately stewing on her guilt, then my friendship risks seeming duplicitous and insincere. As Simon Keller notes, “you want a friend who’s on your side, not one who’s good at faking it.”¹² These worries about insincerity are a specific manifestation of a more general worry for simple evidentialism: that it does not capture the crucial insight that friendship has a robust psychological dimension—that our mental states matter to our friendship just as our acts do.¹³ We can really imagine, for instance, Mateo

10 Rioux focuses on suspension of judgment and disbelief in a friend’s guilt rather than on positive belief in their innocence (“On the Epistemic Costs of Friendship”). The positive view I propose understands the same basic mechanisms as operative in both scenarios. See Soter, “Acceptance and the Ethics of Belief,” 2233–34.

11 For relevant discussion of this feature in the morality of belief more generally, see Basu, “What We Epistemically Owe to Each Other” and “The Morality of Belief I.”

12 Keller, “Friendship and Belief,” 335. See also Hazlett, *A Luxury of the Understanding*, 100–2.

13 Hazlett defends this idea more broadly, writing, “We should not construe morality narrowly so that it pertains only to our actions. . . . We can owe it to [people] to *think* about them in certain ways” (*A Luxury of the Understanding*, 101, emphasis added).

feeling torn about what to believe—feeling like he owes Shelby belief in her innocence despite the bad evidence.

Full-blown partialism targets the cognitive dimension of friendship, taking seriously that we want our friends to be on our side psychologically, even (and perhaps especially) when things do not look good for us. Various philosophers defend the felt tension in such cases, and this instinct even has some empirical support. Some studies have found that participants say that a close friend should, when faced with someone's wrongdoing, believe not only more optimistically about their friend than an acquaintance, but (crucially) more optimistically than they themselves say is rationally permitted by the evidence.¹⁴ In a familiar domain like friendship, it seems reasonable to take such widespread folk and philosophical intuitions seriously, at least as a starting point. Other partialists appeal to more theoretical grounds. For instance, Keller argues that we have a stake in what our friends believe about us, proposing that our friends' beliefs can make a difference to our well-being according to many theories.¹⁵ In a slightly different vein, Sarah Stroud argues that friendship constitutively involves a commitment to believing in the goodness of our friends.¹⁶ Whatever the precise motivation, partialism captures that friendship makes demands not only on how we treat our friends but also on how we *think* about them.

But full-blown partialism faces problems of its own. One kind of problem arises from appreciating that some goods of friendship seem to require evidentially grounded beliefs. Perhaps giving advice about life decisions or showing friends appropriate care requires a clear-eyed view of our friends.¹⁷ Others emphasize the value of honesty and knowing our friends for who they really are: they argue that having fitting emotions towards our friends requires accurate beliefs about them, or they worry that it is inauthentic to believe well about our friends because we think that is what a good friend does rather than because we are attuned to evidence of their positive qualities.¹⁸ These worries each highlight the cost of prescribing irrational belief—especially systematically irrational beliefs—in friendship and thus offer motivations for evidentialism

14 Cusimano and Lombrozo, "Morality Justifies Motivated Reasoning in the Folk Ethics of Belief," 5–13.

15 Keller, "Belief for Someone Else's Sake," 20–24.

16 Stroud, "Epistemic Partiality in Friendship," 501–2.

17 Arpaly and Brinkerhoff, "Why Epistemic Partiality Is Overrated," 43; Kawall, "Friendship and Epistemic Norms," 357; and Dormandy, "Loving Truly," 15–18.

18 Kawall, "Friendship and Epistemic Norms"; Mason, "The Epistemic Demands of Friendship" and "Epistemic Partialism and Taking Our Friends Seriously"; Dormandy, "Loving Truly"; and Crawford, "Believing the Best."

that are rooted in the nature of friendship (rather than in more general epistemological principles).

Perhaps the most poignant worry for full-blown partialism is its prescription that we ought to believe for nonevidential reasons and *against* our evidence. Standard accounts hold that we do not have direct voluntary control over our beliefs; specifically, we cannot choose to believe something unsupported by our evidence directly for practical or moral reasons.¹⁹ Assuming an ought-implies-can constraint on our obligations of friendship, this worry threatens to completely head off the possibility of doxastic obligations of friendship. So although full-blown partialism better captures the cognitive dimension of friendship, it comes with worrisome theoretical baggage and risks overcorrecting from simple evidentialism to the point of devaluing rationality in friendship.

1.3. *Intermediate Positions*

As this debate has unfolded, more nuanced views have been proposed that fall somewhere between these two dialectical poles. I highlight two intermediate routes here. Each holds onto a broadly evidentialist background but moves beyond the “simple” view by considering other potential resources for evidentialists to make sense of partialist intuitions.²⁰ In virtue of this, each also represents a retreat from full-blown partialism; I thus briefly offer reasons to think these strategies might not satisfy someone compelled by partialist motivations. I do not intend these as definitive arguments against these intermediate positions, which I think capture important components of the epistemic landscape of friendship. Rather, I offer these as reasons to think that neither route settles the debate about partialism, and thus, it is worth pursuing further potential accounts.

The first (and commonly appealed to) route points out that even if friendship cannot directly bear on belief, there are myriad “upstream” epistemic practices that shape what we ultimately come to believe—such as how we inquire, gather evidence, verify information, and think through a problem—and friendship

19 E.g., Alston, “The Deontological Conception of Epistemic Justification”; Hieronymi, “Controlling Attitudes”; and Williams, “Deciding to Believe.”

20 Another possible route is to deny the background commitment to evidentialism altogether. This is the route favored by proponents of encroachment, which holds that moral and practical considerations can affect what it is epistemically rational to believe. (For an overview, see Bolinger, “Varieties of Moral Encroachment.”) For present purposes, I set encroachment strategies to the side and accept a background of evidentialism; given its heterodox status, it is a benefit if our solution to the puzzle of partialism does not rely on embracing encroachment. Moreover, for an argument that encroachment fails to resolve the puzzle of partialism, see Rioux, “On the Epistemic Costs of Friendship.”

can clearly impact how we should structure these upstream practices.²¹ This approach amounts to a retreat from full-blown partialism to a weaker position (what Mason calls *indirect partialism* and Arpaly and Brinkerhoff call *partialism-light*), which locates epistemic considerations of friendship in these actions instead. This solution, however, is vulnerable to the worry that no amount of altering upstream practices can guarantee that any *particular* belief state will ultimately be achieved; this is the familiar problem for appeals to “indirect” doxastic manipulations aimed at producing a desired belief state.²² For some, this may be enough: they may think that partialist motivations can be satisfied just via alterations of upstream practices. However, this may not satisfy more committed partialists; some—particularly those motivated by the proposal that it is *especially when* the epistemic chips are down that we most want our friends to believe favorably about us, when others might not—may still think that it matters what belief you end up with.²³

Another route is to argue that we can capture partialist motivations without prescribing anything as strong as belief against the evidence or even irrational belief: perhaps all we owe our friends is a “modest epistemic bias” in their favor—one that does not reach the level of irrationality.²⁴ One way to spell this out is via *epistemic permissivism*, which holds that there is not one unique belief or confidence state that is rationally required by any given body of evidence.²⁵ On this view, friendship can influence which beliefs we adopt from those that fall within the rationally permitted range, thus allowing friendship to be a reason to believe without requiring irrationality or beliefs against the evidence. This reveals two choice points for partialist sympathizers. First, this route is most natural to those independently sympathetic to epistemic permissivism, but it

21 Arpaly and Brinkerhoff, “Why Epistemic Partiality Is Overrated”; Brinkerhoff, “The Cognitive Demands of Friendship”; Goldberg, “Against Epistemic Partiality in Friendship”; and Saint-Croix, “Rumination and Wronging.”

22 See Hieronymi on belief management (“Controlling Attitudes,” 54–56). Arpaly and Brinkerhoff also note that philosophers sometimes seem overly optimistic about how much precise control can be exerted via indirect manipulation (“Why Epistemic Partiality Is Overrated,” 42).

23 For discussion of this idea for doxastic wronging more generally, see Basu, “Morality of Belief 11”: Although appealing to upstream practices very plausibly is part of the morality of belief, it may not be the whole story.

24 Kawall, “Friendship and Epistemic Norms,” 351.

25 Kelly, “Evidence Can Be Permissive”; and Schoenfield, “Permission to Believe.” For discussion, see also Goldberg, “Against Epistemic Partiality in Friendship.” For a broadly permissivist account developed in terms of “epistemic policies” rather than individual beliefs, see Paul and Morton, “Believing in Others.”

may be less attractive to those who favor uniqueness.²⁶ Second, this does not satisfy partialist sympathizers who really do think that reasons of friendship and reasons of epistemic rationality can come into conflict with each other—that, as Keller says, we cannot always meet both the highest standards of epistemic rationality and the highest standards of friendship.²⁷ For some, this may be an acceptable sacrifice to the partialist agenda; but others may want to account for conflicts between norms of epistemic rationality and friendship.²⁸

In what follows, I offer a novel positive alternative that also aims to fall between the two poles of simple evidentialism and full-blown partialism and tries to capture the motivations but avoid the key problems of each. I also show that it can pick up some of the slack that the routes just discussed leave open: it offers an alternative for cases in which the evidence does not plausibly support the desired belief about our friends or in which upstream modifications are not enough. Ultimately, my proposal is compatible with the upstream practices route and modest epistemic bias route; each may capture a different piece of the epistemic landscape of friendship. Given the various choice-points laid out here for partialists, I note once more that the argument I give is a conditional one: *if* one wants to rescue partialist motivations, I offer another potential resource to do so—but I do not here aim to offer new arguments in favor of partialism or to take a stand on how strong of a partialist one should be.

2. A NEW OPTION: ACCEPTANCE

My proposal is this: what we owe our friends in cases like Cheating is not belief but rather *acceptance*. I rely on a specific account of acceptance characterized as suppressing belief's default guiding influence across reasoning, cognition, and action—a notion best understood by its analogy to familiar strategies of emotion regulation. I then discuss how acceptance can help us capture the rich doxastic landscape of friendship in target cases while avoiding the worries plaguing simple evidentialist and full-blown partialist views. A central goal is to motivate that acceptance is a more powerful tool in the ethics of belief than is often appreciated.

Beliefs are states of epistemic confidence concerning the truth or falsity of some proposition: they are our representation of how we take the world to be, given our evidence. But belief states, once formed, do not just sit inert in

26 Uniqueness denies epistemic permissivism and holds that for any body of evidence, there is exactly one rational doxastic attitude. For discussion, see Kopec and Titelbaum, "The Uniqueness Thesis."

27 Keller, "Friendship and Belief," 330.

28 E.g., Vahid, "Friendship and the Grades of Doxastic Partiality."

our minds: they also serve to guide our processes of reasoning, cognition, and action, shaping a wide range of mental and behavioral processes in belief-congruent ways.²⁹ Beliefs are thus, in Michael Bratman's terms, our "default cognitive background": they direct (in conjunction with other mental states) our patterns of thought, goal selection, action, attention, and so on)—and they do it spontaneously and non-inferentially (that is, by default, without our need for conscious oversight).³⁰

Sometimes, however, we do not want to let some belief play its usual role in structuring our deliberation and action. In domains far removed from epistemic partiality, some philosophers propose that in such circumstances we can instead *accept* some alternative—where accepting involves taking some proposition as a premise in practical deliberation and action, even though we do not strictly speaking believe it. In accepting, we thus depart from our default cognitive background of belief: we intervene to prevent belief from playing its characteristic role in guiding reasoning, cognition, and action.³¹ A lawyer, for instance, might have high confidence that her client is guilty but nonetheless accept the client's innocence on professional grounds, committing to a policy of reasoning and acting on the basis of the client's innocence. While belief is often characterized as involuntary and (rationally and psychologically) determined by the evidence, acceptance is proposed to be more clearly under our

29 This is a reasonably well-accepted characterization of the two central functional roles of belief: evidence responsiveness, and guiding inference and action. Notably, if one thinks that it is *only* the output-side guidance function and not an input-side evidence-responsiveness function that characterizes belief (as is arguably the case for some kinds of dispositionalism), then one might not buy the present distinction between belief and acceptance. I consider how such theories of belief interface with the mechanisms discussed here in Soter, "A Defense of Back-End Doxastic Voluntarism." However, specifying the puzzle of partiality in the first place seems to depend on a notion of belief that understands it as centrally an evidence-responsive state: otherwise, we would not face a puzzle about how to think about these apparent nonevidential motivations for belief.

30 Bratman, "Practical Reasoning and Acceptance in a Context," 10. See also Railton, "Reliance, Trust, and Belief," 139. For more discussion of the guidance role and its mechanisms, see Soter, "A Defense of Back-End Doxastic Voluntarism" and "Belief's Guidance Function," 4–5. Specifying exactly how any given belief state shapes these various mechanisms is hard to do with much generality, as it depends on both the content of the belief state and how that interacts with an agent's other mental states, such as their goals, desires, and other beliefs.

31 Acceptance has been discussed by a number of authors (e.g., Bratman, "Practical Reasoning and Acceptance in a Context"; van Fraassen, *Images of Science*; Engel, "Believing, Holding True, and Accepting"; Cohen, "Belief and Acceptance" and *Essay on Belief and Acceptance*). Though there are important differences between these accounts, there is also much overlap. Here, I focus on Bratman-style accounts.

direct voluntary control: we can choose to accept for practical goals, in specific contexts, and in response to nonevidential considerations.

Standard accounts of acceptance are often pitched in these epistemological terms. Elsewhere, I argue that more precisely spelling out the specific cognitive mechanisms that instantiate acceptance offers better insight into its psychological profile.³² I here outline the basics of that proposed account and then turn to the central goal of applying the general account to the specific case of epistemic partiality.

If acceptance involves reasoning, deliberating, planning, thinking, and acting on the basis of something other than what one believes, this requires a monitoring of one's cognition, reasoning, and behavioral processes to identify the various ways in which the unwanted target belief state (or set of beliefs) is active and influencing one's cognition and action, and then intervening to block the usual inferences, actions, patterns of reasoning, thinking, and other downstream effects caused or licensed by the target belief.³³ We can characterize this intervention as a *cognitive gating operation*, through which we prevent the target belief state from having its usual downstream role in cognition, deliberation, and action, and then restructuring the deliberative and inferential landscape accordingly.³⁴ Patterns of thinking, reasoning, and acting that would be licensed (or inhibited) by the belief must now be blocked (or are now permitted) by acceptance.

Why should we think we have such a cognitive gating capacity? I propose that this account inherits theoretical and empirical plausibility via its mechanistic similarity to well-studied emotion regulation strategies; appreciating this can give us a better grip on the cognitive profile of acceptance.³⁵ One prominent class of emotion regulation techniques is *response modulation* (also called *suppression*), which seeks to regulate the characteristic verbal, behavioral, and

32 Soter, "Acceptance and the Ethics of Belief."

33 Existing epistemological notions of acceptance differ in whether they specify that acceptance involves *departing* from one's underlying belief (e.g., Bratman, "Practical Reasoning and Acceptance in a Context") or whether what we reason/act on is what we accept *whether or not we believe it* (e.g., Cohen, "Belief and Acceptance"). I focus on the former because when belief itself guides reasoning/action, there is no explanatory psychological gap to fill: belief is just playing its usual guiding role. For discussion, see Soter, "Acceptance and the Ethics of Belief," 2219–20.

34 Importantly, acceptance does not mean we stop relying on our beliefs *wholesale*. It simply means that there is some *target* belief (or perhaps set of beliefs) whose guiding role we block; of course, we will still be guided by many other beliefs.

35 Soter, "Acceptance and the Ethics of Belief," 2225–31.

cognitive consequences of an elicited emotion.³⁶ Such strategies target the characteristic downstream effects of an activated emotion state—for instance, facial expressions of disgust or vocalizations of fear—without directly targeting the underlying emotion state itself.

I propose that accepting deploys the same cognitive mechanisms at work in emotional response modulation against belief states: acceptance is *doxastic response modulation* (or *doxastic suppression*). With both emotion and belief, an agent seeks to suppress the target mental state by blocking its characteristic cognitive and behavioral effects: these suppression efforts target the consequences of the state rather than directly targeting the underlying state itself. In both domains, we must deploy these regulatory mechanisms for as long as the target state remains intact, and we must remain committed to suppressing it. And crucially, in both the emotional and doxastic domains, response modulation allows an agent to regulate a mental state in response to practical and moral considerations that are not themselves the right kind of reason on which to form the underlying mental state. That is, just as emotion regulation allows us to, for instance, suppress our anger in situations where anger is practically disadvantageous *even if the anger has been fittingly elicited*, so too can doxastic response modulation (i.e., acceptance) allow us to suppress the effects of a belief that is practically or morally undesirable, *even if the belief has been well formed and evidentially justified*. In both cases, we can regulate the underlying states without compromising the proper functioning of the state-elicitation processes.

This account reveals some central psychological characteristics of acceptance, which are familiar in the emotional domain but less appreciated in the doxastic one. First, acceptance is cognitively effortful and demanding on executive processes, of which we have limited capacity: these suppression processes involve effortful inhibition of default patterns of thinking and acting. Second, acceptance turns out to be not a one-off action but rather a temporally extended sequence of specific mental acts. Accepting, in other words, involves committing to gating and restructuring over time: for however long and in whatever contexts an agent is trying to accept (and so long as the underlying belief state remains), she must block and redirect the characteristic downstream effects of her target belief state in cognition and action. This is something the agent can choose to do, but notably, the control profile here is one of effortful regulation over time—a very different control profile than we might initially imagine for

36 Gross, “The Emerging Field of Emotion Regulation” and “Antecedent- and Response-Focused Emotion Regulation”; and McRae, “Cognitive Emotion Regulation.” These are classically contrasted with *antecedent-focused* strategies, which seek to prevent the elicitation or activation of an emotion state in the first place or otherwise intervene upon the generation of the emotion state (akin to “upstream” doxastic intervention).

acceptance. Finally, these response modulation mechanisms can be deployed against any underlying belief or confidence state—that is, one can regulate the default effects of high confidence or belief, or uncertainty. We can thus deploy the acceptance mechanisms discussed here in any context in which the default cognitive and behavioral effects of our underlying belief state are at odds with moral or practical motivations.

2.1. A Question About Doxastic Ontology

This last point—that response modulation can be deployed against any underlying state of confidence—might raise a question about how acceptance fits into broader debates about doxastic ontology, particularly the distinction some have proposed between belief and credence. Though a full treatment of this question is beyond the scope of this article, I here offer some initial thoughts on the matter, given that this distinction has garnered recent attention within the ethics of belief.³⁷

In sections 3 and 4 I will make the case for understanding epistemic partiality in terms of acceptance. Still, readers can presumably already see where I will be going: I will argue that in cases like Cheating, what we should do as friends is not believe against the evidence but rather *accept*: intervene in order to block the underlying belief state from playing its characteristic guiding role. In setting this up, I characterize the agent's belief state as the evidence assessment/state of epistemic confidence that, by default, guides reasoning, cognition, and action.

Recently, some have posited a distinction in doxastic ontology between *credence* as a degreed state of confidence or subjective probability and *belief*, an all-out, categorical, settled state (usually considered one of a tripartite set, along with disbelief and withholding). There is ongoing debate about how best to characterize the relationship between these attitudes, but some suggest that belief plays a distinctive role in guiding inference and action—a role that (even high) mere credence does or should not play.³⁸ Precisely what this distinctive role is varies across accounts. Proposed functions include: guiding reasoning and action, especially in high-stakes contexts; justifying blame; simplifying reasoning by ruling out small error possibilities; and closing inquiry.³⁹ Most relevant for present purposes, some have leveraged the belief/credence distinction to address questions in the ethics of belief—for instance, in explaining why

37 Thanks to an anonymous reviewer for encouraging me to discuss this.

38 For an overview, see Jackson, “The Relationship Between Belief and Credence.”

39 Fantl and McGrath, “Evidence, Pragmatics, and Justification”; Jackson, “How Belief-Credence Dualism Explains Away Pragmatic Encroachment”; Buchak, “Belief, Credence, and Norms”; Staffel, “How Do Beliefs Simplify Reasoning?”; and Friedman, “Inquiry and Belief.”

statistical evidence cannot justify belief about individuals or by highlighting a posited distinctive relationship between beliefs, blame, and inquiry.⁴⁰

Here is the worry: If full belief is distinguished from credence by how it shapes downstream thought and action, and if I say that acceptance involves suppressing the guidance function of an agent's evidence-responsive epistemic assessment, does my proposal just collapse into the belief/credence distinction, and does this therefore render the proposal to be delivered in sections 3 and 4 a redundant move in the debate?

I think not (or at least not obviously). My account of acceptance provides a specific mechanistic story about doxastic regulation and diagnoses a distinctive psychological profile (discussed above and in the subsequent sections). Moreover, my account appeals to independently plausible, well-studied mechanisms of emotion regulation. The debate about credence and belief as two distinct doxastic kinds has largely unfolded from an epistemological perspective; whether and how this distinction can be vindicated cognitive-scientifically, as well as the mechanistic architecture of these two states, are areas of open inquiry in early stages.⁴¹ If the distinction is ultimately vindicated as part of an empirically plausible mental architecture, one possibility is that doxastic response modulation is a mechanistic instantiation of this distinction—for instance, if you think that normally, credence guides reasoning and action, but in cases where we have high credence but want to withhold belief, we need to suppress these default effects. But this might not be the right account. Some dualists (such as Elizabeth Jackson) argue that belief and credence are two fundamentally distinct doxastic states, which form in response to different kinds of reasons (e.g., Lara Buchak argues that naked statistical evidence justifies credence but not belief) and potentially enter into mental computation in different ways (e.g., Julia Staffel's account of credence versus belief in reasoning).⁴²

So, there are two possibilities. The first is that the mechanisms of acceptance are a plausible instantiation of the relationship between credal confidence states and all-out doxastic states, in which case the account I provide here goes far beyond what any existing dualist account says about the psychological profile of all-out belief. The second possibility is that acceptance/doxastic response modulation is just an entirely distinct component of our cognitive economies, in which case working out how these mechanisms interface with both belief

40 Buchak, "Belief, Credence, and Norms"; Moss, "Knowledge and Legal Proof"; and Quanbeck, "Belief, Blame, and Inquiry."

41 See Ballarini, "Credences in Active Reasoning"; Jackson, "The Cognitive Science of Credence"; and Weisberg, "Belief in Psychontology."

42 Jackson, "Why Credences Are Not Beliefs"; Buchak, "Belief, Credence, and Norms"; and Staffel, "How Do Beliefs Simplify Reasoning?"

and credence is an open question for future work. Either way, applying doxastic response modulation to issues of epistemic partiality contributes something new to the landscape of the debate.⁴³ With that laid out, let us turn back to partiality.

3. PARTIALITY AND ACCEPTANCE

We can now offer a new alternative to full-blown partialism and simple evidentialism. The *acceptance view* proposes that in cases where an agent seems pulled in one direction by the evidence and in another by reasons of friendship, what she owes her friend is not belief but acceptance. That is, rather than adopting a belief state inconsistent with the evidence, she has reasons of friendship to regulate the guiding role of a (set of) belief state(s)—to block the characteristic downstream effects of the (well-formed and evidentially justified) belief and commit herself to restructuring her reasoning, thinking, and acting accordingly. Mateo might thus assess that his evidence suggests that Shelby cheated but nevertheless accept that she has not been shown guilty—blocking his assessment of the evidence from guiding his patterns of reasoning, cognition, and action.

The acceptance view carves a middle ground between full-blown partialism and simple evidentialism. It captures that there is something genuinely cognitive and doxastic about the cases at hand, but the obligations involved are not (quite) located in belief formation itself. In section 4, I argue that this approach thus avoids the challenges facing each of the standard views. First, however, let us dive deeper into what the acceptance view suggests about the landscape of cases like Cheating. Specifically, I highlight two key components of the acceptance view: it prescribes cognitive regulation of diverse psychological mechanisms, and it paints a rich diachronic picture of commitment and cognitive work.

3.1. *The Richness of Response Modulation*

Acceptance involves the restructuring of diverse psychological and behavioral processes, blocking belief from guiding these processes as it normally would. Some behavioral manifestations of acceptance are outward and familiar: when Mateo accepts that Shelby did not cheat, he might say supportive things about her to others, give her an important role in the next group project, etc. But this extends inward too: he will also have to prevent himself from reasoning

43 Additionally, in the existing literature on epistemic partiality, it is not clear that authors' discussion of what a friend should believe is specifically meant to be understood in terms of *full/outright belief* (as contrasted with credence by dualists). Rather, the question of interest is whether one's characteristically evidence-responsive doxastic state—which they and I call the agent's belief state—should also be influenced by reasons of friendship.

based on her guilt—for instance, he might prevent himself from concluding that because she cheated, her class grade is going to suffer or that because she cheated on this exam, she might be more likely to cheat on the next one. But once we start to consider these internal dynamics of acceptance, we appreciate that beliefs normally guide a host of cognitive activities: in addition to guiding reasoning, planning, and action, they also guide our patterns of thought, attention, memory, and other various processes. Acceptance involves intervening on these myriad processes. Beyond just avoiding planning and reasoning on the basis of Shelby's guilt, Mateo might also prevent himself from spending time thinking or ruminating about her guilt (e.g., avoid mulling "Why would she do that?"), prevent his worries about her guilt from affecting his patterns of attention (e.g., not focusing on how his text messages about her studying progress went unanswered and instead attending to how determined she was to do well), prevent himself from recalling memories of other times she was guilty (e.g., striving not to dwell on memories of the last time she cheated and instead recalling other times she has surpassed expectations), and so on.

Drawing out these diverse regulatory consequences of acceptance—many of which feel like familiar parts of trying to be a good friend—reveals that acceptance accounts for a variety of ways in which we might think a friend ought to respond cognitively in these kinds of situations. Acceptance thus unifies what might otherwise appear to be a set of independent cognitive duties of friendship and captures the idea that friendship has a significant "internal" or cognitive dimension of maintaining a particular way of thinking about one's friends.

3.2. *Acceptance and Attention*

Let us pause to consider how acceptance interacts with recent proposals that friendship should shape our patterns of *attention* specifically. Anna Brinkerhoff and Catharine Saint-Croix both argue that what we owe our friends is favorable patterns of attention: we owe it to a friend to focus on the good things about them, not to dwell on unfavorable evidence or their bad qualities, and so on.⁴⁴ This approach offers the key insight that attention processes feed into belief in an important way, capturing that there is more to the psychology of friendship than belief itself. The acceptance view agrees with this and even agrees that restructuring patterns of attention is a key part of the landscape—but it goes beyond the (mere) attentional accounts in several ways.

Attention is sometimes talked about (not necessarily by these authors) primarily in its capacity as a process that is "upstream" of belief: what we attend to

44 Brinkerhoff, "The Cognitive Demands of Friendship"; and Saint-Croix, "Rumination and Wronging."

affects belief formation. As with all upstream manipulations, reshaping patterns of attention may indeed affect what beliefs we form—but it cannot be guaranteed to result in any particular belief state. A partialist, then, may still worry about cases in which attention redirection does not result in the favorable belief. Alternatively, we might appreciate attention as something that can be upstream or downstream of belief formation: beliefs (especially about something like a friend's misconduct) affect our patterns of attention in various ways—but we can override that and redirect attention in a way that is more favorable to our friend. This captures a more dynamic interaction between attention and belief and is precisely what the acceptance view agrees with—except the acceptance approach holds that attention is just one of the many cognitive processes that beliefs affect, and thus, we might regulate out of concern for our friends.

The acceptance view, in other words, shares the motivations behind attentional views (and so is friendly to and compatible with them)—that there is something cognitive, but it is not quite belief—but expands on and encompasses these views, highlighting that attention is just one piece of a bigger picture of cognitive regulation. This leaves us on surer footing to tackle the underlying goal: giving a cognitive diagnosis for partialism cases without locating the demands in belief itself. The acceptance view emphasizes that just as there are various things we can do to intervene upstream of belief—including evidence-gathering, attending, rethinking, and so on—there are also things we can do immediately downstream of belief: we are not just at the mercy of our assessments of the evidence once they are formed.

3.3. *Acceptance over Time*

Another core feature of the acceptance view is the rich diachronic story it diagnoses for partiality cases. First, it captures a profile of *cognitive effort* and *commitment*. Acceptance involves a series of effortful mental control actions deployed over time. Suppressing belief's guidance across psychological mechanisms is not simply willed and therefore completed; it involves continuous maintenance for as long as the underlying belief remains intact, and the agent remains committed to regulating it. This takes extended cognitive effort due to engagement of executive control processes—thus requiring a real psychological commitment on behalf of the friend, especially if the belief is frequently activated.

I think this rightly captures how we imagine cases like Cheating playing out: as long as Mateo takes his evidence to speak in favor of Shelby's guilt, his "taking her side" psychologically will involve this psychological effort and commitment. Indeed, accepting can be a psychological burden. In the emotional domain, suppression strategies can over time lead to negative psychological

consequences, including stress, anxiety, and reductions in well-being.⁴⁵ Though it has not been empirically tested, we might predict similar effects in the doxastic domain: acceptance may have psychological costs. But there is nothing mysterious about the idea that friendship can motivate costly actions. We frequently do difficult things for our friends that we might not do for just anyone. Acceptance is another kind of burden we might bear for our friends' sake. This is particularly fitting given what we noted earlier—that the cases that demand thoroughgoing acceptance are not everyday cases but rather particularly challenging ones in which our friends need us “on their side.” It should not be surprising that such cases are psychologically burdensome to us as friends.⁴⁶

That acceptance is constituted by an extended pattern of effortful mental actions also means there are many opportunities for the accepting agent to fail to suppress the downstream effects of their belief. There is a nonzero probability of error (whether of monitoring or of suppression) for each effort to block a downstream effect of belief, and the likelihood of such error increases when an agent's executive processes are engaged elsewhere.⁴⁷ But this possibility of suppression failure also reflects how we might imagine the scenario playing out: we can imagine that when Mateo is overwhelmed by other demanding tasks or distracted (or even drunk!), he might slip up and find himself thinking, acting, or saying something that reveals his underlying assessment of Shelby's guilt. On this picture, accepting can be difficult. But rather than being a problem for the view, I think this gets things exactly right: when the epistemic chips are down, being a good friend can be hard work.

Together, these features reveal something about the action profile of acceptance: it should be understood as an exercise of *self-control*, of regulating behavior and cognition to align with practical commitments. Conceiving of things in this way has normative implications: our moral assessment of an accepting agent should be sensitive to this control profile. For instance, given the difficulty of perfect success in acceptance over long periods of time, an agent may

45 Butler et al., “The Social Consequences of Expressive Suppression”; John and Gross, “Healthy and Unhealthy Emotion Regulation”; and Moore et al., “Are Expressive Suppression and Cognitive Reappraisal Associated with Stress-Related Symptoms?”

46 Of course, this can go too far—there could be cases where the burden is so psychologically demanding that cost of this acceptance becomes too high, and reasons of self-preservation may outweigh reasons of friendship. As I emphasize throughout, my claim is not that one should always accept in the way I propose, only that there are some cases where being a good friend can give us reasons to accept. I also do not claim that acceptance is always a psychological mechanism used for good; clearly, it can be misused in unhealthy interpersonal relationships as well.

47 For relevant discussion, see Sripada, “Addiction and Fallibility” and “The Atoms of Self-Control.”

not be fully culpable for each of these slips, particularly if she is engaged in other cognitively demanding tasks that leave fewer resources available for upholding acceptance.⁴⁸ Of course, a friend might still feel upset about such slips, and the agent might in some sense be answerable for them. Nonetheless, it seems we ought to recognize that an agent who commits to accepting something on behalf of her friend is doing something morally serious, even if—given her cognitive limitations—she does not do so perfectly.⁴⁹ This point also highlights something significant about the broader project here: better understanding the cognitive dynamics of the cases of interest can reveal normative complexity that may not have been antecedently apparent.

Finally, reflecting on sustained acceptance highlights that the interaction between acceptance and belief is highly dynamic: in particular, acceptance might eventually shape belief. Restructuring thought, reasoning, attention, and action as guided by belief at one point in time may well lead to changes in how an agent acquires and understands information and evidence, which inferences she does (not) draw, and so on—in a way that may ultimately influence her underlying belief states. So although on the acceptance view, the stated goal is not to alter a particular belief state (on pain of being self-undermining), acceptance over time may well eventually alter our beliefs—and the line between acceptance and belief may blur over time, especially if patterns of acceptance become learned and habitual. (In section 5, I consider whether this is a problem for the account.)

4. AVOIDING OBJECTIONS TO THE STANDARD VIEWS

The preceding section aimed to describe the novel features of the doxastic landscape of friendship drawn out by the acceptance view. Let us now revisit some of the problems for full-blown partialism and simple evidentialism raised in section 1. As we return to these, it is useful to keep in mind exactly where the acceptance view aims to intervene in the dialectic. First, it aims to capture the strong partialist intuition that there can be substantial conflicts between the norms of friendship and epistemic rationality, and these are not exhausted merely by modifications in upstream epistemic practices or plausibly captured

48 For discussion in the context of addiction, see Sripada, “Addiction and Fallibility.”

49 Further, our friend might recognize this: that you are putting in the (cognitive) work for their sake. This point stands in interesting tension with the methodology used in some accounts of doxastic wronging (e.g., Basu and Schroeder, “Doxastic Wroning”), which appeal to feeling *owed an apology* for unfavorable beliefs. Perhaps if we recognize our friend’s evidential situation, the story about owing apology is not so simple—we might recognize that it is hard to block one’s beliefs in this way and that a friend’s doing so is a sign of their commitment to us.

by a “modest epistemic bias.” We are setting aside views that deny the need to capture this strong partialist intuition regarding genuine conflicts between what our evidence supports and what we feel we should believe as friends. Second, however, the acceptance view wants to hold on to a broadly evidentialist orthodoxy that although there seems to be a real conflict here, we cannot just believe directly in response to nonepistemic reasons of friendship. Thus, the acceptance view ultimately is an evidentialist view (or at least, is compatible with evidentialism, hence my setting aside encroachment views); but it aims to do more to satisfy those with strong partialist intuitions than the simple evidentialist position traditionally captures. It does so by pointing out that there is a lot of internal psychological regulation we can do on behalf of our friends via acceptance—even if this regulation falls short of directly modifying belief formation.

4.1. *Challenges for Simple Evidentialism*

Recall that the simple evidentialist pole of the partiality debate holds that reasons of friendship cannot bear directly on belief; they can bear only on action. With this as the starting point, one way simple evidentialists could try to account for motivations of friendship is to say that as a friend, we should *act as if* we believe something about our friends—but we should not *actually* believe it. (This strategy is perhaps not actually defended by anyone in print, but it is a clear possibility in the space of the debate.) This approach brings up two related worries. First, it elides a key psychological dimension of friendship; second (and partly in virtue of the first), if it locates the influence of friendship only on outward action, it risks prescribing a kind of pretense towards our friends that is insincere or deceptive.⁵⁰

Regarding the former, we have seen that the acceptance view diagnoses a rich psychological landscape for partiality cases. Without positing direct obligations to form beliefs against the evidence, the acceptance view nonetheless acknowledges the importance of beliefs in our cognitive economies and prescribes robust regulation of those beliefs, thus avoiding simple evidentialism’s naive suggestion that only our outward behavior needs modifying.

But the insincerity concern remains worrisome for the acceptance view. Is acceptance as I defend it so different from the “acting as if” approach for

50 Views by authors like Mason (“The Epistemic Demands of Friendship” and “Epistemic Partialism and Taking Our Friends Seriously”), Crawford (“Believing the Best”), and Dormandy (“Loving Truly”) avoid this worry by saying that there is a robust psychological dimension to friendship, but it is not a *partial* one—it is instead manifested via our accurate knowledge of our friends. They thus avoid the worry about prescribing pretense—but they also deny the strong partialist intuition about the possibility of *conflict* between epistemic and friendship motivations, which I here aim to capture.

which I criticize simple evidentialism? This question reveals that the “acting as if” locution is often vague: it underdescribes an agent’s psychological profile. Distinguishing between different precisifications of this notion reveals that worries about insincerity have more bite against some versions than others.

There is a possible “thin” version of acting as if that is located entirely in the agent’s external behavior and includes nothing “in the head.” This is the version I presented for “simple” evidentialists at the start, casting them as a foil to psychologically focused partialists. On this thin profile, the agent continues to let belief (e.g., in her friend’s guilt) drive her inner life—intervening only right before she hits the point of external behavior or speech. This profile involves a dramatic mismatch between the agent’s internal life and external actions. Alternatively, we can conceive of acting as if in a “thick” sense, in which the agent intervenes far earlier in the process, on her cognitive as well as behavioral dynamics, redirecting her patterns of thought, attention, reasoning, and so on. On this version, there is a lot going on psychologically—she is just not altering the underlying *belief* itself. But this thick notion of acting as if just looks like what I am calling acceptance.

Let us thus distinguish acceptance (the thick psychological profile) from “mere” acting as if (the thin, entirely behavioral profile). Worries about insincerity and deception really hit home against “mere” acting as if, but they have less force against acceptance because the former, but not the latter, involves a split between the agent’s inner and outer lives. Mere acting as if is brittle, both normatively and practically: the agent seems insincere both due to the mismatch between how she thinks and how she acts and because she always seems just one missed action away from revealing how she is really thinking about her friend. In contrast, precisely what is distinctive and attractive about acceptance is its thoroughgoingness: accepting involves a real commitment to restructuring one’s internal mental life. Acceptance captures the cognitive faith in our friends that partialists are after: we are taking a risk, on behalf of our friend, of thoroughly altering how we act *and think* about our friend despite unfavorable evidence—and so broadcasting a genuine commitment to our friend. This involves significant cognitive work—just not work that is (directly) beliefaltering. Thus, the rich psychological story of acceptance helps alleviate worries about sincerity: an agent can recognize that her evidence looks bad for her friend but nonetheless decide that because of friendship, she is not going to let that assessment of the evidence dictate how she thinks (and acts).⁵¹

51 We might even grant that the agent can be transparent with her friend about this. We can imagine Mateo saying to Shelby, “Look, we both know the evidence looks bad. But you’re my friend, so I’m committed to not letting that assessment of the evidence influence how I think and act.” I have intentionally (though somewhat awkwardly) avoided using the

4.2. Challenges for Full-Blown Partialism

The objections against full-blown partialism fall into two main categories: those that target the friendship part of the thesis, and those that target the belief part. Let us begin with the latter.

4.2.1. Objections from Belief

The most obvious objection to full-blown partialism is the problem of doxastic control. On standard philosophical accounts, we lack direct voluntary control over our beliefs, which are thought to be rationally and directly responsive only to evidence—that is, we cannot choose to believe something unsupported by our evidence directly on the basis of practical or moral reasons that we take to have no bearing on the truth of the claim in question.⁵² This orthodoxy threatens to entirely undercut full-blown partialism: we do not owe epistemic partiality because we *cannot* believe against the evidence for reasons of friendship.⁵³ Importantly, some authors talk about partiality in terms of what we *owe* our friends and what friendship *demand*s of us: for instance, Keller writes that “the tendency to treat us sympathetically [through their beliefs] is not just one that we think likely to be manifested in our friends, it is one that we can *want* them to manifest.”⁵⁴ At least some versions of partialism are thus framed as telling us what we ought to do to be good friends (more on this shortly)—making the ought-implies-can-style worries pressing.

Of course, even if we cannot believe *directly* in response to nonevidential reasons of friendship, no one denies that we have various kinds of *indirect* control over our beliefs: we can shape what beliefs we come to have through what evidence we gather, who we spend our time with, how we inquire, and so on. This takes us back to the route of upstream epistemic practices discussed in section 1.3. But this may not really satisfy partialists: as Arpaly and Brinkerhoff point out, philosophers are sometimes a bit too happy to throw around the idea that we can cultivate specific mental states in ourselves—this is actually quite

words ‘acceptance’ and ‘belief’ here because I do not think that the colloquial/folk uses of those terms are well defined enough to capture the precise way in which we are carving things up here.

52 Classically, among many others, Alston, “The Deontological Conception of Epistemic Justification”; and Williams, “Deciding to Believe.” For helpful discussion, see Hieronymi, “Controlling Attitudes,” who argues that belief is responsive specifically to reasons that *bear on the question of whether p* (50–52).

53 Arpaly and Brinkerhoff, “Why Epistemic Partiality Is Overrated,” 40–41; and Goldberg, “Against Epistemic Partiality in Friendship,” 223on17.

54 Keller, “Friendship and Belief,” 338.

difficult to do with much precision or reliability.⁵⁵ This worry is particularly bad for belief: if we alter our epistemic practices with the stated goal of trying to bring about a particular belief in ourselves that we do not take to be supported by our evidence, this process risks becoming self-undermining.⁵⁶

The acceptance view straightforwardly meets the challenge of doxastic control. Acceptance provides us with an account of how friends ought to respond when evidential reasons for belief and reasons of friendship conflict, without forcing us either to deny the orthodox view that we cannot choose to believe for practical/moral reasons or to retreat to indirect methods of belief manipulation that may be at best unreliable and at worst self-undermining. Acceptance takes seriously the limits of our control over belief formation but highlights the substantive regulatory control we have over the role that beliefs play in our cognitive economies.

In section 1.3, we considered a possible intermediate position between simple evidentialism and full-blown partialism: the argument that perhaps friendship does not actually generate *conflicts* with our evidence but instead merely requires a “modest epistemic bias.”⁵⁷ Though this is a plausible piece of the epistemic landscape of friendship, I noted that it may not satisfy those who want to capture the idea that friendship and epistemic rationality can be at genuine odds with each other.⁵⁸ The problem with this approach is that a partialist can always continue presenting cases that further stack the evidential deck against the friend. Absent a principled argument as to why reasons of friendship and epistemic rationality can never conflict, our goal should not be to explain why some particular belief is actually rational but rather should be to make sense of the cases in which friendship and epistemic rationality conflict—whatever those cases may be.⁵⁹ Acceptance explains how we can respond doxastically in cases where the evidence, by stipulation, does not rationally or psychologically permit a positive belief in our friends. In doing so, it may help alleviate the temptation to try to explain away all cases of apparent evidential irrationality in terms of, for instance, permissive standards of belief—because one (perhaps implicit) motivation for appealing to permissivism may precisely

55 Arpaly and Brinkerhoff, “Why Epistemic Partiality Is Overrated,” 42. See also Hieronymi, “Controlling Attitudes.”

56 Williams, “Deciding to Believe,” 148–49.

57 Kawall, “Friendship and Epistemic Norms.”

58 For a recent characterization that frames the debate in these terms, see Woodcock, “If Epistemic Partialism Is True, Don’t Tell Your Friends.”

59 In general, I am skeptical of attempts to show that (even permissivist accounts of) evidentialism can *always* explain away conflicts between morality and epistemic rationality. See Traldi, “Uncoordinated Norms of Belief.”

be to avoid running into worries about doxastic control. So although permissive epistemic standards, perhaps coupled with altered upstream epistemic practices, may well account for many cases where we believe differently about our friends than a disinterested observer would, we need not (and, I think, should not) assume that there can *never* be dilemmas that pull us between epistemic norms and moral ones. Acceptance—which is responsive to a wider range of reasons than the merely evidential—gives us the resources to explain what we owe our friends in those cases where what seems doxastically demanded by friendship also seems genuinely outside of what it is rationally permissible to believe based on our evidence.⁶⁰

These arguments that the acceptance view can avoid worries of doxastic control or prescribed belief against the evidence draw our attention to another possible response to such worries—and a bigger issue in the background of the partialism debate. This is the question of what kind of normativity is at stake for partialist epistemic claims. Are partialists aiming for a *prescriptively* normative story that tells us, in a guidance-giving way, what we ought to do as friends? Or are they instead after a merely *evaluatively* normative claim that delivers assessments of how it would be *good* for friends to believe? Though I flagged above that some authors (arguably Keller and Stroud) talk in terms of what friendship can *demand* of us or what we can *owe* them, which suggests a prescriptive account, others explicitly offer merely evaluative assessments of friendship-beliefs.⁶¹ This route may be less susceptible to worries about doxastic control, as it can plausibly be good for us to do things even if they are not under our voluntary control.

Whether to offer a prescriptive versus merely evaluative account represents another choice point for partialists. I take it to be an advantage of the acceptance view that it offers the possibility of a prescriptive story, appealing to the kind of (mental) action we can choose to deploy on behalf of our friends. Still, this is compatible with a merely evaluative assessment that sometimes our tendency to accept on behalf of our friends is good (at least from the perspective of friendship) without the stronger commitment that there are cases where we are *obligated* to do this. But for those sympathetic to the goal of explaining what an agent can decide to do when they find a conflict between what their evidence

60 There are big-picture questions in the background about rationality conditions for acceptance, including whether they should be understood in terms of merely practical or also partly epistemic rationality. This is an important theoretical question for the view, which I do not try to settle here. The key point is just that appealing to acceptance does not run into the *same* kinds of issues of irrationality that prescribing belief formation against the evidence does.

61 Crawford, “Believing the Best”; and Dormandy, “Loving Truly.”

supports and the needs of their friend, acceptance offers us an actionable route towards navigating such situations.⁶²

4.2.2. Objections from Friendship

The second group of challenges for the partialist view includes worries that we *should not* believe partially about our friends because some important goods of friendship require our beliefs to be rational and evidentially grounded.

There are several versions of this worry. First, in some contexts, having an irrationally partial view of our friends might interfere with fulfilling certain roles of friendship, such as giving good advice when friends face big life choices or risky decisions.⁶³ Sound advice-giving depends on having sufficiently accurate views of our friends; in this respect, irrationally favorable beliefs could make us *worse* friends. But all this shows is that we *sometimes* need honesty from our friends' beliefs. This does not rule out that in other contexts, we need partial beliefs—such as when our friends need support—rather than frank advice. Yet this maneuver poses its own problem for partialism: on standard accounts, beliefs are characteristically stable across contexts (barring changes in one's evidence). Thus, a friend who succeeds in adopting an irrationally favorable belief about my chances of business success in one context cannot simply discard that belief in a context where evidentially justified beliefs become important.

But there is a deeper version of the worry about rational belief in friendship: we want our friends to know and love us based on *who we really are*, not some fictionalized, idealized version of us.⁶⁴ And rational honesty may matter not only for assessments of our friends' flaws: Lindsay Crawford argues that even when we believe *well* about our friends, we should do so not because we think

62 Another possible question about normativity concerns whether the normativity at stake here is epistemic, pragmatic, or a *sui generis* notion from friendship. See Stroud, "Epistemic Partiality in Friendship," 502. This depends on one's views about these domains of normativity; I do not try to settle that here. Even whether the acceptance-mechanisms described here should be understood as epistemic or practical is a complicated matter; if one limits epistemic normativity to merely the norms that guide belief formation, acceptance could be categorized as practical, but one could reasonably also think that the question of whether a belief state plays its guiding role relates to questions of epistemic normativity.

63 Arpaly and Brinkerhoff, "Why Epistemic Partiality Is Overrated," 43; and Kawall, "Friendship and Epistemic Norms," 360. Both worry about such cases.

64 Kawall points out that if a friendship is based on systematic illusion about someone, we should worry that the friendship is flawed and "not a love of the friend herself, with her actual character and qualities" ("Friendship and Epistemic Norms," 361). Similarly, Mason proposes a Murdochian account of friendship that centers around knowledge of a friend's true character ("The Epistemic Demands of Friendship"). Both (but especially Mason) suggest that systematically irrational beliefs undermine the legitimacy of friendship, preventing one from loving and relating to *the person themselves*.

that is how a good friend should believe but because we think those beliefs are rationally warranted—because we are attuned to the good qualities of our friends that justify those beliefs.⁶⁵ Crawford holds that this is necessary for authentic engagement with our friends.

Together, these worries charge full-blown partialism with undervaluing rational belief in friendship and—if partialists try to limit the scope of their claim to particular contexts—with being unable to explain how belief can be context dependent.

The acceptance view fares better on both fronts. It can handle the worries about honesty and authenticity related to knowing and loving our friends for who they really are, because the view recommends no *systematic irrationality* towards our friends. It makes no demands on the normal functioning of belief-forming mechanisms and so is compatible with thinking that having an accurate assessment of our friends is generally a component of friendship. The view merely holds that in some (exceptional) cases, we have reasons of friendship to prevent our beliefs from playing their characteristic guiding role in cognition, reasoning, and action—committing ourselves to a supportive stance towards our friends.⁶⁶ Further, the acceptance view holds up to Crawford-style authenticity worries: although it would perhaps be inauthentic to believe well only on the basis of reasons of friendship as a general matter, in the fraught cases at hand, reasons of friendship are precisely what we seem to be responding to. It is *because* of their friendship that Mateo accepts Shelby's innocence. This does not seem inauthentic; rather, acceptance explains how we can be responsive to reasons of friendship as such in these conflict cases.

The acceptance view also makes room for the idea that some contexts of friendship involve rational belief, and others involve partiality. Unlike belief, which is characteristically context stable, acceptance can be context dependent. Thus, explaining partiality in terms of acceptance gives us an important kind of discretion: we can accept when doing so would be beneficial for the friendship (e.g., when a friend needs someone to have their back) but not when rational belief is more valuable (e.g., when they need clear-eyed advice or an honest intervention). Keller too notes the need for this context sensitivity, writing, "Good friends treat each other differently under different circumstances, and a good friend often has the skill of being able to discern and respond to her friend's needs, as they are and as they change ... and the same goes for belief

65 Crawford, "Believing the Best."

66 On the role of acceptance in proleptic trust, see Frost-Arnold, "The Cognitive Attitude of Rational Trust." Frost-Arnold draws on Bratman's notion of acceptance in trust, where belief is not rationally permitted; the cognitive profile developed here works well with her account.

formation.”⁶⁷ This feature, puzzling on the standard belief framing, is no problem for acceptance.

But perhaps there is a lingering worry about authenticity stemming from the lawyer example used above to introduce acceptance. Are the lawyer and the friend supposed to be the same—and if so, does this suggest another kind of inauthenticity from the friend? After all, we do not think of a lawyer accepting their client’s innocence as being particularly authentic.

This worry allows us to spell out another nuance of the view: that the scope of the acceptance mechanism can vary across two key dimensions. The first is the range of contexts in which an agent deploys belief regulation mechanisms. A lawyer perhaps does so in professional or court contexts but seems limited to those—we would not think she should continue accepting when she is out socializing with her friends after the case has concluded. In contrast, friendship may motivate accepting across a much wider range of contexts. Second, there is the question of which characteristic effects of belief an agent regulates; someone can be more or less comprehensive on this dimension. The lawyer needs to change her patterns of courtroom speech and behavior but does not seem to owe her client a commitment to reasoning based on his innocence more broadly. In contrast, precisely what is attractive about acceptance in friendship is its thoroughgoingness: that it involves the regulation of a wide range of psychological responses. The comparison of the friend and lawyer holds a broader lesson: acceptance mechanisms can be deployed in a variety of ways, and how we *should* deploy them depends on the normative considerations in play. If we are accepting to fulfill the normative demands of a particular relationship, the scope of the acceptance mechanisms we should deploy depend on the nature of that relationship and what it takes to meet its demands. And friendship—or so partialists think—is a relational context in which the demands of acceptance can be particularly thorough.

But relational demands may not be the only considerations in play. Another worry for partialists concerns weighing reasons of friendship against other kinds of moral reasons. Consider, instead of cheating, an accusation of sexual assault: even if someone might in some sense be a good *friend* in accepting someone’s innocence, they risk serious moral harm to the victim in doing so. This highlights that reasons of friendship are just one of the many considerations that govern acceptance. In deciding whether to accept, we must consider not only the needs of our friend but also various other kinds of moral factors, including

67 Keller, “Belief for Someone Else’s Sake,” 28. Warman agrees, writing, “Friendship requires nuance. As friends we must master the fine art of recognizing when our friends need the benefit of the doubt, and when they need us to be especially cautious about the possibility of error” (“Epistemic Partiality and the Nature of Friendship,” 386).

considerations of justice to others involved. In some cases (such as Cheating, which is relatively mild), reasons of friendship may dominate. In others, other moral concerns may outweigh friendship.⁶⁸ Figuring out when and whether acceptance is all-things-considered appropriate is part of the complex project of both friendship and morality. My goal here is not to defend whether acceptance is ultimately the right choice in any particular kind of case. Rather, as noted at the outset, I aim to show that *if* there are times when considerations of friendship seem to conflict with epistemic rationality, acceptance is a resource to explain how we should and potentially actually do handle these scenarios.

5. TWO WORRIES ABOUT COHERENCE

Let us briefly consider two worries regarding the coherence of an accepting agent.

5.1. *Agential Incoherence*

My argument contends that an agent can take a body of evidence to support a belief that *p* but also suppress that belief's guiding role. One might worry that this puts the agent in an oddly divided state: she concurrently takes herself to have sufficient reason to believe *p* (e.g., Shelby cheated) and also sufficient reason to act and reason on the basis of not-*p* (e.g., the case has not been made that Shelby cheated). Does this kind of internal division somehow threaten one's agential coherence?

The parallel to emotion regulation again helps us here. Emotions carry assessments of situations (even beliefs, on some views) and guide our downstream cognition and action. Yet it is uncontroversial that an agent can rationally appraise a situation as warranting an emotion—say, anger—but also recognize that expressing that anger outwardly or letting it structure her inner reasoning and thoughts is prudentially (or even morally) inappropriate. In other words, an agent can recognize anger as a *fitting* response to some situation—that it is “angery” —and also recognize that she ought nevertheless to resist letting anger structure her reasoning, cognition, and action—because the anger is undesirable for reasons other than those relating to the fittingness or rationality of anger elicitation.⁶⁹

This is a deeply familiar situation. A child might do something that rationally elicits their parent's anger; the parent might nonetheless have strong reason

68 Baker (in “Trust and Rationality”) and Rioux (in “On the Epistemic Costs of Friendship”) agree that obligations of partial belief in friendship are *prima facie*.

69 D'Arms and Jacobson, “The Moralistic Fallacy.”

to prevent that anger from affecting her cognition, reasoning, and action. A politician's opponent might say something angersome, yet the politician has strong reason to suppress her anger. And so on. The claim that an agent who prevents her anger from guiding her reasoning, thinking, and action as it would if left unchecked is somehow agentially incoherent simply lacks bite. Indeed, it seems a sign of her status as a globally integrated agent that she has the capacity to decide whether to let her emotions guide her or whether to override them.

My proposal is that emotion and belief are structurally and normatively analogous. We very often have strong reason to allow our beliefs to guide us; so too we often have strong rational reason to let our emotions guide us, especially when we have no reason to doubt that our affective systems are well attuned. But we do not and should not always leave our emotions—or our beliefs—unchecked. Just as we can appraise a situation as rationally warranting an emotion but nonetheless (for a different set of reasons) choose to suppress that emotion, so too can we assess a situation as rationally warranting a belief state but nonetheless (for a different set of reasons) choose to suppress that belief from having its usual cognitive and behavioral effects. This is not agentially suspect in the case of emotions; neither, I propose, is it for beliefs. In fact, our capacity to regulate these mental states actually expands our psychological agency rather than undermining it. And indeed, capturing this feeling of conflict is apt in the cases we are concerned with: these are precisely meant to be hard cases in which we feel pulled between competing significant values and considerations (e.g., rationality and accuracy versus friendship).

Is there still some remaining inauthenticity at stake here? Perhaps, insofar as there is a split between an agent's evidence assessment and how she decides to think and act. But the crucial thought is that the agent's doxastic intervention may well be more reflective of her values and commitment to her friendship than her mere evidence assessment is: she is, because of a commitment to her friend, intervening on the part of her belief process, over which she has control. This, I suggest, is not objectionably inauthentic or incoherent in quite the same way that the original concern worries about—because her commitment to her friend is entirely sincere. And this, I think, goes a long way towards getting us what we might want out of epistemic partialism.

5.2. *Diachronic Incoherence*

I suggest that acceptance avoids the partialist pitfall of directly prescribing belief against the evidence. But I also acknowledge that an agent who systematically redirects her patterns of thinking, reasoning, and acting may ultimately end up with different beliefs than she otherwise would have, as that restructuring affects the evidence she gets and how she understands it, what kinds

of inferences she does and does not draw, what questions she thinks through, and so on. This opens up a worry for the acceptance view regarding *emergent diachronic irrationality*: the accepting agent risks acquiring warped epistemic states down the line, ending up with beliefs that are less accurate than they could have been and that are inconsistent with her original beliefs—as a result of accepting for reasons of friendship rather than for truth-conducive reasons. In other words, even if acceptance does not prescribe immediate irrationality as full-blown partialism does, do its prescriptions nonetheless ultimately lead to irrationality, incoherence, or epistemic suboptimization down the line?

Such changes in underlying beliefs are indeed a possible consequence of sustained acceptance, as noted in section 3.2. However, I do not think this ultimately poses a serious problem for the acceptance view. First, there is the question of whether acceptance actually leads to belief change in any given case. It might or might not; this depends on how the agent's acceptance interacts with the evidence she has and continues to acquire. It is entirely plausible that if the evidential circumstances remain similar, the underlying belief itself will not actually change much.

But even when acceptance does lead to belief change, there are two key features about the way in which it does that help alleviate concern. First, the complaint against the full-blown partialist prescription of belief against the evidence is not a complaint about irrationality as such; rather, the problem is that it prescribes belief in response to reasons of friendship without offering a solution to the problem of doxastic control. Beliefs resulting from sustained acceptance (to the extent that they occur) avoid this issue because they are brought about via normal doxastic mechanisms—one does not accept with the stated goal of altering one's beliefs; rather, those changes result down the line from changes in the evidence that one has and how one understands that evidence. Because of this, we still avoid more the most pressing worries that face full-blown partialism about direct doxastic control and self-undermining in attempted belief manipulation.

Second, presumably, the beliefs likely to result from accepting on behalf of our friends are ones that are more favorable towards them. This may sometimes result in epistemic errors that we would not otherwise make—like if Mateo comes to really believe that Shelby is innocent when she is not. But other times, it may lead us towards accurate beliefs that we would not otherwise have—like if it turns out that Shelby really is innocent, despite the evidence being stacked against her. Notice that the pattern of errors this will bias an agent towards is consistent with partialist motivations: we might think that when it comes to our friends and loved ones, certain kinds of errors are more desirable than others—e.g., it might be better from the perspective of friendship

to believe a friend is innocent when she is guilty, than the reverse.⁷⁰ Though someone with no partialist sympathies at all might not find this an attractive result, to the extent that the acceptance view is trying to capture the intuitions motivating partialism, this pattern seems to be a virtue rather than a problem. So perhaps the accepting agent ends up with different beliefs and a different epistemic landscape than she otherwise would have. But whether this kind of incoherence is really a problem for the acceptance view is not obvious; to the extent that it is, it depends on the adjudication of more general debates about the importance of coherence, which will have to be met on their own terms.⁷¹

6. CONCLUSION

I have defended a novel position in the ongoing debate over epistemic partiality—namely, that in cases like Cheating, friends should *accept*. That is, they should regulate the characteristic guiding role of an unwanted belief across cognition, reasoning, and action. I have proposed that this offers a middle ground between the two traditional camps: simple evidentialism, which denies that friendship can give us reasons to believe (and thus which relegates any obligations of friendship to the domain of outward behavior), and full-blown partialism, which holds that we sometimes owe our friends belief against the evidence. I have suggested that the acceptance view captures the underlying motivations of each, highlighting the cognitive demands of friendship without eliding the value of rational belief, while being less susceptible to the most pressing objections against the standard views.⁷² Whether acceptance is indeed enough to satisfy a

70 For a defense of this kind of idea for humanity more broadly, see Preston-Roedder, “Faith in Humanity.”

71 See, e.g., Fogal, “Rational Requirements and the Primacy of Pressure”; Kolodny, “Why Be Rational?” and “Why Be Disposed to Be Coherent?”; Lasonen-Aarnio, “Enkrasia or Evidentialism?”; and Worsnip, “The Conflict of Evidence and Coherence” and “Making Space for the Normativity of Coherence.”

72 One could worry that acceptance still will not *really* satisfy partialists because partialists hold that our friends really want from us *genuine belief*, not belief-adjacent acceptance. We might imagine Shelby saying to Mateo, “I don’t just want you to *accept* that I’m innocent. I want you to *actually believe* it.” One tricky component of this is that ‘belief’ and ‘acceptance’ as used in this article are technical terms that may not map neatly onto folk language. (See also note 51 above.) For discussion of how these mechanisms relate to different conceptions of belief, see also Soter, “Belief’s Guidance Function” and “A Defense of Back-End Doxastic Voluntarism.” Let us redescribe this scenario without using either word. We could imagine Mateo responding, “Look, we both know the evidence looks bad [*showing his belief: assessment of the epistemic situation*], but you’re my friend and because of that, I’m committed to not thinking that you’re guilty even in spite of that evidence [*describing his acceptance*] and not letting this evidence shape how I think about or treat

committed partialist may depend on what precisely they think the psychological component of friendship demands; my goal has simply been to argue that acceptance can offer a much richer picture than we may have assumed, and I leave it to partialist sympathizers to decide whether they think this goes deep enough.

An additional goal of this article has been to show that acceptance can be a powerful tool for addressing questions in the ethics of belief. To that end, even if the acceptance view as a position within the epistemic partiality debate is not free from problems or limitations, I hope to have shown that when we dig down into the psychological profile of acceptance, we can see that it has more resources to diagnose the landscape of these cases than has often been appreciated.⁷³ Moreover, many of these insights come into focus specifically when we consider its cognitive dynamics—this angle, in particular, is a key novel feature of the present approach. That said, one possible upshot of this analysis is that closer consideration of the psychological demands of acceptance could lead to doubt regarding whether accepting really *is* a good thing to do in the end. As I have noted throughout, my aim here is thoroughly conditional: to offer a psychologically plausible account of an intuitive component of friendship but not to defend that we really do owe this to our friends. If it turns out that closer inspection of these psychological dynamics ends up leading us to doubt that acceptance is a healthy component of friendship in the end, then we have still made progress in the debate in a way that we would not have without thorough consideration of the psychological landscape.

A final theoretical advantage of the account defended here is that acceptance provides a unified framework for reframing diverse problems in the ethics of belief. Acceptance is not special to friendship; it is an independently plausible, general-purpose cognitive regulation strategy that we can deploy when we have moral or practical reason not to let belief play its default guiding role.⁷⁴ Accep-

you at all." This, I think, strikes us as fair in the situation. Admittedly, it might not get us the strongest possible version of full-blown partialism, but as I say throughout, my goal is to offer an account of what we can actually do when we find ourselves in such unfavorable evidential situations regarding our friends.

- 73 Two notable exceptions to the tendency to dismiss acceptance as a solution are Jørgensen (published as Jørgensen Bolinger), "The Rational Impermissibility of Accepting (Some) Racial Generalizations"; and Begby, *Prejudice*. Jørgensen appeals to acceptance in the context of racial profiling-style generalizations, though the notion of acceptance she relies on is different than the one I develop here. Begby indicates his shared friendliness to acceptance as an approach to a range of puzzles in the ethics of belief. Though he does not develop the psychological profile in the way I do here, much about our views is compatible.
- 74 Similarly, Goldberg argues that a virtue of his account of how values can affect our upstream epistemic practices is its general theoretical soundness ("Against Epistemic Partiality in Friendship," 2235).

tance perhaps could help us understand a variety of tricky cases—for example, when an agent seems to have an evidentially justified but morally unsettling statistical belief or when accurate beliefs about one's chances of success are psychologically detrimental. That acceptance might help us in various issues with worries about doxastic control and practical reasons for belief at their heart is another consideration in favor of affording it a role in the epistemic partiality debate. Further, this account of acceptance strives to be psychologically plausible, aiming to describe processes that we *actually do undertake*. My hope is that acceptance thus described feels like a familiar part of our cognitive lives and of our friendships—that it captures a way of relating to each other that we really do deploy and that in at least some cases, it is good (from the perspective of friendship) that we do so.⁷⁵

One of the central upshots of the acceptance view is that friendship can demand significant cognitive work. A commitment of partialists is that it is an essential component of being a good friend that we can think about friends in ways that we might not think about others. Though we may not be able to simply will away unwanted beliefs (or emotions) regarding our friends, we *can* exercise significant control over the ways in which these mental states structure our cognitive lives—and, at least sometimes, being a good friend involves making this effort. On the one hand, the fact that friendship requires this cognitive work is something that few likely want to deny. But on the other hand, working through this idea reveals something deeply important about the nature of our relationships to others: the needs and good of other people can make significant moral demands on how we structure our cognitive lives.⁷⁶

York University
lksoter@yorku.ca

- 75 Warman expresses a similar hope for capturing the psychology of friendship, noting that a theory of friendship should capture that we are often “doing friendship right” (“Epistemic Partiality and the Nature of Friendship,” 384): our normative theories of friendship should cohere reasonably well with the kinds of things we actually seem to do for our friends.
- 76 I am grateful to Chandra Sripada, Peter Railton, Renée Jørgensen, Maegan Fairchild, Zach Barnett, Juan S. Piñeros Glasscock, and several anonymous referees for their feedback on various drafts of this article; and to Susan Gelman, Ethan Kross, Angela Sun, Gabrielle Kerbel, Jake Lehrle-Fry, Max Lewis, John Monteleone, Matt Stichter, Paul Rezkalla, John Doris, Carissa Phillips-Garrett, Mark Herman, and Michael Milhim for further discussion. Versions of this article were presented at Duke University, University of Michigan (including the Graduate Student Working Group), California Institute of Technology, York University, Washington and Lee University, Arkansas State University, the 2022 Munich Graduate Conference in Ethics, the 2022 Rocky Mountain Ethics Congress, the 2022 Central New York Moral Psychology Workshop, and the 2023 Pacific American Philosophical

REFERENCES

- Alston, William P. "The Deontological Conception of Epistemic Justification." *Philosophical Perspectives* 2 (1988): 257–99.
- Arpaly, Nomy, and Anna Brinkerhoff. "Why Epistemic Partiality Is Overrated." *Philosophical Topics* 46, no. 1 (2018): 37–51.
- Baker, Judith. "Trust and Rationality." *Pacific Philosophical Quarterly* 68, no. 1 (1987): 1–13.
- Ballarini, Cristina. "Credences in Active Reasoning." Unpublished manuscript, 2025.
- Basu, Rima. "The Morality of Belief I: How Beliefs Wrong." *Philosophy Compass* 18, no. 7 (2023): e12934.
- . "The Morality of Belief II: Three Challenges and an Extension." *Philosophy Compass* 18, no. 7 (2023): e12935.
- . "What We Epistemically Owe to Each Other." *Philosophical Studies* 176, no. 4 (2019): 915–31.
- Basu, Rima, and Mark Schroeder. "Doxastic Wronging." In *Pragmatic Encroachment in Epistemology*, edited by Brian Kim and Matthew McGrath. Routledge Press, 2019.
- Begby, Endre. *Prejudice: A Study in Non-Ideal Epistemology*. Oxford University Press, 2021.
- Bratman, Michael E. "Practical Reasoning and Acceptance in a Context." *Mind* 101, no. 401 (1992): 1–15.
- Brinkerhoff, Anna. "The Cognitive Demands of Friendship." *Pacific Philosophical Quarterly* 104, no. 1 (2023): 101–23.
- Buchak, Lara. "Belief, Credence, and Norms." *Philosophical Studies* 169, no. 2 (2014): 285–311.
- Butler, Emily A., Boris Egloff, Frank H. Wilhelm, Nancy C. Smith, Elizabeth A. Erickson, and James J. Gross. "The Social Consequences of Expressive Suppression." *Emotion* 3, no. 1 (2003): 48–67.
- Cohen, L. Jonathan. "Belief and Acceptance." *Mind* 98, no. 391 (1989): 367–89.
- . *Essay on Belief and Acceptance*. Oxford University Press, 1995.
- Crawford, Lindsay. "Believing the Best: On Doxastic Partiality in Friendship." *Synthese* 196, no. 4 (2019): 1575–93.
- Cusimano, Corey, and Tania Lombrozo. "Morality Justifies Motivated Reasoning in the Folk Ethics of Belief." *Cognition* 209 (2021): 104513.
- D'Arms, Justin, and Daniel Jacobson. "The Moralistic Fallacy: On the 'Appro-

- priateness' of Emotions." *Philosophy and Phenomenological Research* 61, no. 1 (2000): 65–90.
- Dormandy, Katherine. "Loving Truly: An Epistemic Approach to the Doxastic Norms of Love." *Synthese* 200, no. 3 (2022): 1–23.
- Engel, Pascal. "Believing, Holding True, and Accepting." *Philosophical Explorations* 1, no. 2 (1998): 140–51.
- Enoch, David. "What's Wrong with Paternalism: Autonomy, Belief, and Action." *Proceedings of the Aristotelian Society* 116, no. 1 (2016): 21–48.
- Fantl, Jeremy, and Matthew McGrath. "Evidence, Pragmatics, and Justification." *Philosophical Review* 111, no. 1 (2002): 67–94.
- Feldman, Richard, and Earl Conee. "Evidentialism." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 48, no. 1 (1985): 15–34.
- Fogal, Daniel. "Rational Requirements and the Primacy of Pressure." *Mind* 129, no. 516 (2020): 1033–70.
- Friedman, Jane. "Inquiry and Belief." *Noûs* 53, no. 2 (2019): 296–315.
- Frost-Arnold, Karen. "The Cognitive Attitude of Rational Trust." *Synthese* 191, no. 9 (2014): 1957–74.
- Gertken, Jan, and Benjamin Kiesewetter. "The Right and the Wrong Kind of Reasons." *Philosophy Compass* 12, no. 5 (2017): e12412.
- Goldberg, Sanford C. "Against Epistemic Partiality in Friendship: Value-Reflecting Reasons." *Philosophical Studies* 176, no. 8 (2019): 2221–42.
- Gross, James J. "Antecedent- and Response-Focused Emotion Regulation: Divergent Consequences for Experience, Expression, and Physiology." *Journal of Personality and Social Psychology* 74, no. 1 (1998): 224–37.
- . "The Emerging Field of Emotion Regulation: An Integrative Review." *Review of General Psychology* 2, no. 3 (1998): 271–99.
- Hazlett, Allan. *A Luxury of the Understanding: On the Value of True Belief*. Oxford University Press, 2013.
- Hieronymi, Pamela. "Controlling Attitudes." *Pacific Philosophical Quarterly* 87, no. 1 (2006): 45–74.
- . "The Wrong Kind of Reason." *Journal of Philosophy* 102, no. 9 (2005): 437–57.
- Jackson, Elizabeth. "The Cognitive Science of Credence." In *The Oxford Handbook of the Cognitive Science of Belief*, edited by Neil van Leewuen and Tania Lombrozo. Oxford University Press, forthcoming.
- . "How Belief-Credence Dualism Explains Away Pragmatic Encroachment." *Philosophical Quarterly* 69, no. 276 (2019): 511–33.
- . "The Relationship Between Belief and Credence." *Philosophy Compass* 15, no. 6 (2020): e12668.

- . “Why Credences Are Not Beliefs.” *Australasian Journal of Philosophy* 100, no. 2 (2022): 360–70.
- John, Oliver P., and James J. Gross. “Healthy and Unhealthy Emotion Regulation: Personality Processes, Individual Differences, and Life Span Development.” *Journal of Personality* 72, no. 6 (2004): 1301–34.
- Jorgensen Bolinger, Renée. “The Rational Impermissibility of Accepting (Some) Racial Generalizations.” *Synthese* 197, no. 6 (2020): 2415–31.
- . “Varieties of Moral Encroachment.” *Philosophical Perspectives* 34, no. 1 (2020): 5–26.
- Kawall, Jason. “Friendship and Epistemic Norms.” *Philosophical Studies* 165, no. 2 (2013): 349–70.
- Keller, Simon. “Belief for Someone Else’s Sake.” *Philosophical Topics* 46, no. 1 (2018): 19–36.
- . “Friendship and Belief.” *Philosophical Papers* 33, no. 3 (2004): 329–51.
- Kelly, Thomas. “Evidence Can Be Permissive.” In *Contemporary Debates in Epistemology*, edited by Matthew Steup, John Turri, and Ernest Sosa. Wiley-Blackwell, 2013.
- Kolodny, Niko. “Why Be Disposed to Be Coherent?” *Ethics* 118, no. 3 (2008): 437–63.
- . “Why Be Rational?” *Mind* 114, no. 455 (2005): 509–63.
- Kopec, Matthew, and Michael G. Titelbaum. “The Uniqueness Thesis.” *Philosophy Compass* 11, no. 4 (2016): 189–200.
- Lasonen-Aarnio, Maria. “Enkrasia or Evidentialism? Learning to Love Mismatch.” *Philosophical Studies* 177, no. 3 (2020): 597–632.
- Mason, Cathy. “The Epistemic Demands of Friendship: Friendship as Inherently Knowledge-Involving.” *Synthese* 199, no. 1 (2021): 2439–55.
- . “Epistemic Partialism and Taking Our Friends Seriously.” *American Philosophical Quarterly* 61, no. 3 (2024): 233–43.
- McRae, Kateri. “Cognitive Emotion Regulation: A Review of Theory and Scientific Findings.” *Current Opinion in Behavioral Sciences* 10 (2016): 119–24.
- Moore, Sally A., Lori A. Zoellner, and Niklas Mollenholt. “Are Expressive Suppression and Cognitive Reappraisal Associated with Stress-Related Symptoms?” *Behaviour Research and Therapy* 46, no. 9 (2008): 993–1000.
- Moss, Sarah. “Knowledge and Legal Proof.” In *Oxford Studies in Epistemology*, vol. 7, edited by Tamar Szabó Gendler, John Hawthorne, and Julianne Chung. Oxford University Press, 2022.
- Paul, Sarah K., and Jennifer M. Morton. “Believing in Others.” *Philosophical Topics* 46, no. 1 (2018): 75–96.
- Preston-Roedder, Ryan. “Faith in Humanity.” *Philosophy and Phenomenological Research* 87, no. 3 (2013): 664–87.

- Quanbeck, Z. "Belief, Blame, and Inquiry: A Defense of Doxastic Wronging." *Philosophical Studies* 180, no. 10 (2023): 2955–75.
- Railton, Peter. "Reliance, Trust, and Belief." *Inquiry* 57, no. 1 (2014): 122–50.
- Rioux, Catherine. "On the Epistemic Costs of Friendship: Against the Encroachment View." *Episteme* 20, no. 2 (2021): 247–64.
- Saint-Croix, Catharine. "Rumination and Wronging: The Role of Attention in Epistemic Morality." *Episteme* 19, no. 4 (2022): 491–514.
- Schoenfield, Miriam. "Permission to Believe: Why Permissivism Is True and What It Tells Us About Irrelevant Influences on Belief." *Noûs* 48, no. 2 (2014): 193–218.
- Shah, Nishi. "A New Argument for Evidentialism." *Philosophical Quarterly* 56, no. 225 (2006): 481–98.
- Soter, Laura K. "Acceptance and the Ethics of Belief." *Philosophical Studies* 180, no. 8 (2023): 2213–43.
- . "Belief's Guidance Function: Mechanisms, Control, and Categorization on the Output-Side of Belief." In *The Oxford Handbook of the Cognitive Science of Belief*, edited by Neil Van Leewen and Tania Lombrozo. Oxford University Press, forthcoming.
- . "A Defense of Back-End Doxastic Voluntarism." *Noûs* 59, no. 1 (2025): 112–39.
- Sripada, Chandra. "Addiction and Fallibility." *Journal of Philosophy* 115, no. 11 (2018): 569–87.
- . "The Atoms of Self-Control." *Noûs* 55, no. 4 (2021): 800–24.
- Staffel, Julia. "How Do Beliefs Simplify Reasoning?" *Noûs* 53, no. 4 (2019): 937–62.
- Stroud, Sarah. "Epistemic Partiality in Friendship." *Ethics* 116, no. 3 (2006): 498–524.
- Traldi, Oliver. "Uncoordinated Norms of Belief." *Australasian Journal of Philosophy* 101, no. 3 (2022): 1–13.
- Vahid, Hamid. "Friendship and the Grades of Doxastic Partiality." *Theoria* 90, no. 1 (2024): 122–33.
- van Fraassen, Bas C. *Images of Science: Essays on Realism and Empiricism*. University of Chicago Press, 1985.
- Warman, Jack. "Epistemic Partiality and the Nature of Friendship." *Ethical Theory and Moral Practice* 27, no. 3 (2024): 371–88.
- Weisberg, Jonathan. "Belief in Psyontology." *Philosopher's Imprint* 20, no. 11 (2020): 1–27.
- Williams, Bernard. "Deciding to Believe." In *Problems of the Self*. Cambridge University Press, 1973.
- Woodcock, Scott. "If Epistemic Partialism Is True, Don't Tell Your Friends."

Analysis 84, no. 3 (2024): 566–75.

Worsnip, Alex. “The Conflict of Evidence and Coherence.” *Philosophy and Phenomenological Research* 96, no. 1 (2018): 3–44.

———. “Making Space for the Normativity of Coherence.” *Noûs* 56, no. 2 (2022): 393–415.

ELICITORY STRUCTURAL POWER AND AGENTIAL POWER AN OUTLINE AND DEFENSE

Arash Abizadeh

THE THESIS that some intentional agents have structural power clashes with two powerful intuitions: that the power of agents operates only by way of their intentional actions and that it consists in the capacity to play some causal role in effecting outcomes. These intuitions suggest that notions of passive and noncausal power illicitly conflate power with prospects for success. As Brian Barry argues, cases in which “desired outcomes occur with no intervention” by agents do not evidence their power because one should “distinguish between those who do well by exercising power and those who are the passive beneficiaries of the activities of others.”¹ Similarly, Keith Dowding asserts that being the passive beneficiary of social structures is a matter of “systematic luck,” not power.²

Yet these intuitions clash with familiar phenomena. We often recognize that it is precisely because some are so powerful that they can satisfy their wishes without having to lift so much as a finger. Picture a mafia godfather whose henchmen, with nary a word from him, anticipate and carry out what they take to be his wishes. His power is the envy of every mafioso reduced to monitoring the details himself—and reduced to pleading, cajoling, and bribing others to secure their service. Or think of a charismatic prophet who so inspires admirers that they anticipate and serve her every wish even before she herself becomes conscious of them. The power she wields over people is the dream of every

1 Barry, “The Uses of ‘Power,’” 348.

2 Dowding, *Rational Choice and Political Power*. Terence Ball suggests that exercising power conceptually implies causally efficacious intentional action (“Power, Causation, and Explanation,” 211). For the view that power must operate by way of actions, see Laswell and Kaplan, *Power and Society*, xiv; Simon, *Models of Man*, 11; Dahl, “Power,” 410; Goldman, “Toward a Social Theory of Power,” 225–26; and the overview in Ball, “Models of Power.” For the view that power implies causation, see Simon, *Models of Man*, 11; Dahl, “Power,” 410; March, “An Introduction to the Theory and Measurement of Influence,” 437; Nagel, *The Descriptive Analysis of Power*, 11; and Isaac, *Power and Marxist Theory*, 74.

would-be spiritual leader. Or consider the structural position of men in our patriarchal societies: a man can often satisfy his preferences or aims in a way a woman cannot, sometimes thanks not to any intentional actions of his own but to society's gender structure, regulated by norms and epistemic frames that dispose others to defer to men's preferences or aims. Something similar is true of white people in European and European-settler societies.

My agenda here is fourfold: to show that active or *agential* power can be efficaciously exercised by way of intentional actions even when the outcomes would obtain without those actions; to defend a notion of *elicitory* power as a type of nonintentional or passive power by which agents elicit welcome outcomes, but not by way of their intentional actions; to defend a notion of *structural* power as a type of elicitory power that agents have in virtue of their positions in social structures; and to defend a *noncausal* category of power—whether agential or elicitory—at stake when outcomes obtain in virtue of an agent's power but without the agent having played any causal role in producing them.³

Many actors at the bottom of our societies' power hierarchies intuitively recognize structural modes of power, which they frequently articulate by drawing on the vocabulary of "privilege."⁴ I argue, however, that these phenomena are often properly understood as instances in which privileged actors have and wield social *power*. This matters because privilege implies inegalitarian, hierarchical power relations, and while many instances of structural power—those grounded in gender or racial hierarchies, for example—fit this mold, other instances may be widely distributed, equally shared, or reciprocal. The recognition of nondecisive, elicitory and structural, and noncausal categories of power is significant because it serves the practical purpose of identifying those over whom it would be useful to wield power; the moral purpose of assigning responsibility and blame; and the evaluative purpose of critically appraising social arrangements in light of their distribution of power.

Readers be forewarned: precisely because moral responsibility is often premised on causal responsibility, I deploy a series of assassin cases below to tease out and clarify latent intuitions about the latter by appealing to clearer judgments about the former. But a drawback is that assassin-style cases can give the misleading impression that power is inherently power over others and hence hierarchical, or that *power-over* is inherently evil. Neither is the case. The concept of power defended here is not a moralized one: having or wielding power is not inherently wrong or evil, and even power-over may be welcome or beneficial to those over whom it is wielded.

3 On active versus passive power in the sense I employ here, see Morris, *Power*.

4 Harris, "Whiteness as Property."

I have three terminological notes and one methodological comment before proceeding. First, the concept of power concerning us here is not the concept to which metaphysicians refer in phrases such as “the causal powers of entities,” which we might call *entity power*. We are here exclusively concerned with the power of agents *qua* intentional agents, which we might label *agent power*—the core notion of which is the capacity to obtain what one might want. Both agential and elicitory power are species of agent power in this sense. I therefore use the term ‘agent’ strictly in the action theory sense of an *intentional* agent, not, as some metaphysicians do, to refer to any entity that produces changes in other entities’ kind, structure, causal powers, or intrinsic properties (as in the phrase ‘causal agent’). Concomitantly, by elicitory or “passive” power I mean the agent power that intentional agents have independently of their intentional actions, not, as these metaphysicians do, the power of an entity to suffer fundamental changes.⁵

Second, the concept of structural power defended here is distinct from three other similarly labelled concepts. ‘Structural power’ is sometimes used to refer to the power of social structures.⁶ Social structures might be said to have causal power, for example, insofar as they constitute what Fred Dretske calls “structuring causes,” i.e., “background conditions that enable one thing to cause the other” were the former to occur.⁷ (Dretske contrasts structuring causes, which establish potential causal pathways, to “triggering causes,” which are events that cause the first element of a causal process to *occur*—now.) I set this alternative usage aside, not because the concept is unimportant but because my concern here—whether in the case of agential or structural power—is with the social power of agents *qua* intentional agents. By analogy with agent power, I call the power of structures *structure power*. ‘Structural power’ is also sometimes used to refer to the power of agents *over* social structures, to create or shape them.⁸ By analogy with structuring causes, I call this *structuring power*. Finally,

5 Marmodoro, “Aristotelian Powers at Work”; and Kuykendall, “In Defense of the Agent and Patient Distinction.” The question of whether intentional agents have elicitory power does not turn on whether there is a viable ontological distinction between “active” and “passive” causal powers (and between causal agents and patients) in this metaphysical sense. Cf. Heil, *The Universe as We Find It*, 74; and Ingthorsson, “Causal Production as Interaction.” Nor by the causal power of agents do I mean, as defenders of “agent causation” do in metaphysical debates about free will, agents’ capacity directly to cause action-triggering intentions. See O’Connor, *Persons and Causes*.

6 Hayward, *De-Facing Power* and “On Structural Power”; Elder-Vass, *Causal Power of Social Structures*; Forst, *Normativity and Power*; and Hasan, “Republicanism and Structural Domination.” Cf. Dowding, *Rational Choice and Political Power*, 8–9.

7 Dretske, *Explaining Behavior*, 42.

8 Strange, *States and Markets*; and Roy, *Socializing Capital*.

some use the term to refer simply to the power that an agent has in virtue of their position in a social structure.⁹ This is a comprehensible sense of structural power—call it the “broad” sense. But my focus is on a more restricted sense because we are here concerned with *social* power, and almost all social power is ultimately structural in this broad sense.¹⁰ I take *social structures* to be constituted by social relations to some extent stabilized by a set of background expectations, rules, norms, schemas, or practices enacted by agents operating within a certain *habitus*—agents whom Thomas Wartenberg calls “peripheral agents” and Nicholas Vrousalis calls “regulators.”¹¹ Few instances of social power do not depend on social structures—and hence are not structural to some degree—in this sense. The broad sense is therefore largely redundant. (Even the power to lift boulders is, in a social context, a power one has partly in virtue of the fact that other agents are not disposed to prevent one from lifting boulders—because of the property regime, for example.) Thus two conditions must be satisfied for power to be structural in the strict sense at stake here: it must be in virtue of the agent’s social-structural position, yes, but it must also be elicitory power, i.e., not operate by way of one’s intentional actions.

Third, I do not mean the verb in phrases such as ‘to effect an outcome’ to be a synonym for *causing* an outcome: to *cause* is to *affect* what happens, but to *effect* what happens is, as I use the term, to realize, or, as the *Oxford English Dictionary* puts it, “accomplish” a preference, intention, objective, or plan.¹² Causally affecting an outcome is neither sufficient nor necessary for accomplishing one’s intentions: one may cause unwelcome outcomes; and, as I argue, one may accomplish one’s intentions or objectives without causing them. I restrict the terms ‘effecting’ and ‘exercising power’, moreover, to the case of *agential* power. By contrast, when outcomes are obtained by way of an agent’s *passive* power (and so not by way of their intentional actions), I say the agent *elicits* (rather than effects) the outcome—which is why I label this type of power *elicitory*. And I use the expression ‘to *obtain* an outcome’ indifferently between actively effecting or passively eliciting. Finally, I use the verb ‘wield’—as in the phrase ‘wielding power’—indifferently between the intentional, active and the nonintentional, passive modes of power: one may “wield” agential or elicitory power, but one

9 Marsh, “Interest Group Activity and Structural Power”; Isaac, *Power and Marxist Theory*; Gädeke, “Does a Mugger Dominate?”; and Vrousalis, “The Capitalist Cage.” Cf. Stone, “Systemic Power in Community Decision-Making”; Ward, “Structural Power”; and Haslanger, “Oppressions Racial and Other.”

10 Isaac, *Power and Marxist Theory*.

11 Bourdieu, *La domination masculine*; Wartenberg, *The Forms of Power*; and Vrousalis, “The Capitalist Cage.”

12 Morriss, *Power*, 29–30.

may “exercise” agential power only. (In using ‘wield’ in this way, to include a passive sense, I am resurrecting its obsolete meanings, which, according to the *OED*, include “to have the . . . advantage of” and to “accomplish” or “achieve, attain.”)

These are of course stipulations on my part—terms of art used in the service of articulating a theory of social power. So my defense of that theory does not rely on appealing to readers’ linguistic intuitions about such terms. Rather, my method is to present paradigmatic cases that recognizably instantiate agents’ social power, to guide our considered judgments about the concept; and to interpret those cases in light of what I take to be the core notion of agent power. To this core we now turn.

1. THE WELCOME, INTENTIONAL ACTION, AND LINKAGE TESTS

Consider a victim who would not have been robbed had he not displayed his wallet.¹³ The victim’s action of pulling out his wallet to give alms is one of the robbery’s antecedent causes: but for his actions, the mugging would not have occurred. Yet although the victim helped cause the outcome, he did not *exercise power* (over the thief, for example) to effect it. Similarly, consider an aspiring leader who, having commanded her would-be subjects to march, is served their loud refusal. Although the aspiring leader has *caused* them to do something—loudly refuse—she has *failed* to efficaciously exercise power: she did not effect the refusal.¹⁴

Why? Because the notion of power concerning us here is not the metaphysical concept of entity power (*qua* causal entity) but rather agent power (*qua* intentional agent). At the heart of this latter notion is the *capacity to obtain what one might want*. The distinctive feature of efficaciously wielding agent power—whether agential or elicitory—is that the outcome must be welcome to the agent: she must favor it in some sense. The hapless victim has not *ex post exercised* power to effect his own mugging because the agent-outcome relation fails what I call the *welcome test*. To be sure, given the presence of muggers, *ex ante* the agent does *possess* such power, and on other occasions, he might exercise it. Imagine the “victim” were a police officer conducting an undercover sting on a known mugger: then the officer would indeed be exercising his power to get the mugger to mug him.

The welcome test comes in both an *ex post* and an *ex ante* version. When we retrospectively inquire whether an agent has efficaciously *wielded* power to obtain an outcome, we apply an *ex post* test, to wit: Was the outcome favored by

13 Morriss, *Power*, 29.

14 Ball, “Power, Causation, and Explanation,” 205.

the agent? The *ex ante* version, by contrast, pertains to whether an agent prospectively *has* the power to obtain a potential outcome. We therefore apply a counterfactual test, to wit: If the agent were to favor the outcome, would it obtain?

The welcome test applies to all forms of agent power: it articulates the core notion underlying both agential and elicitory power. It might be objected that this construal fails to account for *unwelcome* power. The objection rests on a mistake. Consider John Stuart Mill, who lamented the arbitrary power Victorian legal structures gave him over his wife.¹⁵ Although *possessing* this power was unwelcome, its *nature* consisted in a capacity to obtain outcomes concerning his wife should he favor them. Mill may have deplored being able to get his way—and perhaps refused to exercise a power he deemed unjust—but his power over his wife consisted in being able to do so. Of course, in one respect, he was powerless: he could not free her from conjugal subjection without systemic legal reform. But the fact that he was powerless to relinquish his power over her does not imply he lacked it.

If the welcome test is common to both agential and elicitory power, the former's distinctive feature is its intrinsic link to exercising intentional *agency*. This has two aspects. First, agency is manifested in intentional actions, i.e., actions constituted by an intention-in-action.¹⁶ Second, intentional action is responsive to one's intentional states, i.e., subjective mental states with representational content. Exercising agential power therefore requires satisfying two corresponding conditions beyond the welcome test. First, the outcome must obtain by way of one's intentional actions. Call this the *intentional action test*.

Second, the fact that the outcome is welcome must be appropriately linked to one's intention-in-action. In particular, one's intention-in-action, which renders intentional action responsive to one's intentional states, must be explained by one's favoring the outcome (i.e., by the favoring attitude satisfying the welcome test). This *linkage test* connects the welcome test to the intentional action test and hence to the exercise of agency. The link is tightest, of course, when one intends the outcome, i.e., when the intentional object of one's intention-in-action is the same as the object of the favorable attitude that satisfies the welcome test and explains one's intention. However, to restrict agential power to intended effects, as many propose, would be to construe exercising agency too narrowly.¹⁷ That the agent intends the outcome is *sufficient* for satisfying the linkage test, but it is not necessary, for two reasons.

15 Mill, *The Subjection of Women*.

16 Searle, *Making the Social World*, 33.

17 Russell, *Power*, 23; Ball, "Power, Causation, and Explanation"; Debnam, "Nondecisions and Power"; Wrong, *Power*; Searle, *Making the Social World*; and Forst, *Normativity and Power*.

First, agents sometimes produce unintended outcomes, as by-products of their actions, which they nevertheless favor in ways sufficiently linked to their exercise of agency to count as their having effected them. Consider the fanciful case of Bumbling Master, inspired by Plautus's comedies.¹⁸ The domestic Lar has set things up so that whenever Master intentionally acts to treat Slave badly, he unintentionally treats her well (and vice versa). So far, so bumbling: whatever Master does, he never satisfies the welcome test and so never exercises his agential power. But now imagine that Master has caught on to Lar's setup and so begins deciding strategically: when he prefers to treat Slave badly, he decides to treat her well—and thus adopts and acts on the intention to treat her well (and vice versa). The welcome test is now satisfied, insofar as he effects the outcomes he prefers. The intentional action test is also satisfied: he effects his preferred outcomes by way of his actions. The outcomes, however, are *unintended*: he intends to treat Slave badly but ends up treating her well. Nevertheless, his intention-in-action is sufficiently linked to his preference: the former is directly explained by the latter.

Consider now a less fanciful case.

Predatory Movie Mogul: A movie mogul holds tremendous power over the careers of women hoping to star in his films. He uses this power to coerce them into sex. His predatory actions have numerous unintended by-products. First, women in the industry, aware of his willingness to abuse his power, come to fear him, resulting in a culture of deference to his artistic judgments on set. Second, after years of apparent impunity, the predator is arrested, convicted, and imprisoned. Although both by-products were unintended, when acting, he had a preference for a culture of subservience, but not for incarceration.

The mogul clearly effected the rapes in virtue of exercising his agential power: the outcomes were intentionally caused by him. But although he caused his own imprisonment, he did not effect it: the outcome was neither intended nor favored by him in acting. By contrast, it seems the culture of subservience was effected by his exercise of agential power because, although unintended, in acting, he favored the particular outcome in question—he preferred it—in a way sufficiently linked to his actual intention-in-action. How so? One way in which his preference (for a culture of subservience) might have been linked to his intention (to subordinate sexually) would have been if that preference had *directly* explained his intention. This would establish a sufficient link, to be sure, but the typical way in which a preference explains an intention is by bringing

18 I owe this example to Niko Kolodny.

about an intention with the same object as one's preference—which simply means the outcome was intended. (So, for example, he might have intended to coerce sex partly in order to foster a culture of subservience, in which case he also intended to foster that culture as well.) But in the mogul's case, his preference (for a culture of subservience) is sufficiently linked to his intention (to sexually exploit) in a weaker way: the former is an instance of a more general attitude—for example, a preference for subordinating women—which in turn helps explain his actual intention-in-action. Thus, the mogul exercises his power in effecting the culture of subservience: while his intention (to sexually exploit) is not caused by his preference for this culture, this preference is an instantiation of a more general preference (for subordinating women) that does cause the intention. Moreover, a hypothetical intention to effect a culture of subservience is consistent with his actual intention. His favoring attitude explains his intention only in this weaker, more extended sense.

Second, people often exercise agency via actions whose intentional objects are not the particular outcomes extrinsic to the actions themselves. (The action event itself is the intrinsic consequence of acting.) Consider a prime minister facing uncertain circumstances.¹⁹ She does not reliably conjecture any of the potential extrinsic consequences of her available courses of action and consequently, in acting, does not intend any particular extrinsic consequence. But this does not imply she cannot exercise her agency or effect outcomes. If the linkage test were reduced to intended effects, then it would follow, absurdly, that the prime minister has no agential power in virtue of her high office to effect outcomes extrinsic to her action.

If the prime minister thought that despite the unpredictability of her actions' consequences, she could nevertheless increase the likelihood of satisfying her general objectives and plans, her actions may have been caused and constituted by a *general* intention to fulfill those objectives; if so, then the intended effects test might suffice to explain her agential power and its efficacious exercise, without resorting to a weaker linkage condition. The "general intention" retort, however, is often unavailable. Perhaps she was acting out of habit, unreflectively applying a rule of thumb in uncertain circumstances, or out of a sense of duty to formal procedural considerations independently of outcomes.²⁰ In that case, even if the extrinsic, downstream consequences of her action fulfilled her objectives or plans, they would not have fulfilled any of her intentions when acting—even a general intention. Yet it seems that even here, she might have efficaciously exercised power.

19 White, "Power and Intention," 751–52.

20 White, "Power and Intention," 756.

We can gain analytical clarity by assuming that from among the menu of possible extrinsic outcomes, the prime minister is in fact indifferent between them. How could she have exercised power in effecting the particular set of outcomes that ensue despite not favoring them? Imagine that in acting, the prime minister favors the circumstance that whichever extrinsic outcomes ensue do so *by way of her agency*—that is, she favors acting as she does and favors the efficacy of her agential power in so acting, i.e., she favors her intentional actions counting as an instance of her exercising power to effect those outcomes. Call this favoring her own *power efficacy*. The upshot is that although she may not favor the particular extrinsic consequences of her actions in and of themselves—whatever they may be—she does favor them indirectly insofar as they are effected by way of her exercise of intentional agency—that is, she favors the set of *overall* comprehensive outcomes: the combination of her actions' extrinsic consequences, her action itself (the intrinsic consequence of her action!), and her power efficacy in effecting those particular outcomes. Then her favoring attitude towards the overall set of consequences, including the extrinsic consequences, is appropriately linked to her exercise of agency because it explains her actual intention-in-action. We can therefore explain why, when the prime minister acts using the powers of her office, she *ex post* successfully effects the extrinsic outcomes—despite not intending or even favoring them in particular.

This analysis can similarly explain the following case.

Devout Tyrant: A fanatically devout tyrant rules over his kingdom with the sole objective of carrying out the directives of his religious adviser. The tyrant has no other objectives or cares and so is completely indifferent to his subjects' plight. He consequently is not motivated by and, in acting as ruler, does not hold in view the effects of his actions on his subjects. Yet his subjects' well-being wholly turns on his actions.

Despite the facts that none of his actions' extrinsic outcomes are intended by him—they are all by-products!—and that he does not favor any of them in particular, he nevertheless exercises immense power in effecting them. The outcomes are appropriately linked to his agency because two conditions are met: first, given that he favors his own actions and their power efficacy, he favors the *overall* set of consequences of his intentional actions (which include his actions' intrinsic consequence—namely, the actions themselves and his power efficacy in effecting the extrinsic consequences); and, second, favoring his own power efficacy is appropriately linked to his actual intention-in-action because it helps explain it. Like the prime minister, the extrinsic consequences are part of a set he favors, and he favors the power efficacy of his actions. The same

could be said of a high-ranking civil servant who cares and intends to ensure only that she follow formally correct procedure in effecting outcomes and who consequently undertakes numerous actions that significantly affect many lives.

Indeed, even if there were some extrinsic consequences the prime minister, tyrant, or civil servant *disfavors*—but not enough to outweigh the extent to which they favor acting efficaciously—they would nevertheless effect even the regrettable by-products. Members of the upper classes whose actions foreseeably help perpetuate lower-class misery exercise power in effecting those outcomes, even if those outcomes are unintended by-products that they would prefer to avoid, given that their reticence is outweighed, all things considered, by the considerations prompting them to continue as they do. Or consider a lieutenant who unintentionally causes subordinates to adopt his outlook but deems such an outcome regrettable because it deprives him of advice from diverse viewpoints.²¹ He might nevertheless be exercising his power insofar as this regrettable outcome is part of an overall package caused by his actions and towards which he is favorable, and he favors his actions being a mode of exercising power over his subordinates. Or consider ruthless industrial “disrupters” whose aim is just to muck around and see what happens: even if the extrinsic results are not in themselves welcome to them in particular, those results may nevertheless satisfy both the welcome and linkage tests insofar as they result from the exercise of their agential power.

To sum up, according to the linkage test for agential power, an agent’s favoring the outcome (which favoring satisfies the welcome test) must be appropriately linked to the intention-in-action that constitutes the action (by way of which the outcome is effected, satisfying the intentional action test). This linkage test can be met in two ways. First, if the agent favors the *particular* outcome in question, then the test is met if her favoring attitude either itself directly explains her intention-in-action or is an instance of a more general favoring attitude that explains her intention (so that her favoring attitude explains the intention in an extended sense). The case in which the agent straightforwardly intends the outcome is only one way of instantiating this. Second, if the agent favors an *overall* set of outcomes of which the particular outcome is a part and favors her own power efficacy in effecting this set of outcomes, and her favoring attitude explains her intention-in-action, then the linkage test is met for the set of outcomes—including the particular outcome in question, even if, on its own, the agent disfavors it.

21 Partridge, “Some Notes on the Concept of Power,” 114.

2. AGENTIAL CAUSAL POWER

A paradigmatic case of having agential power and efficaciously exercising it is provided by the following case.

Unique Assassin: A victim has taken refuge in a location inaccessible to anyone except one assassin. If not for the unique assassin, the victim would live until old age. But the assassin accesses the location and shoots, intentionally killing the victim.

Ex ante, the unique assassin *has* the agential power to kill the victim (and is the only one with the power to do so herself). *Ex post*, the assassin has efficaciously *exercised* her power to effect the victim's killing. I wish to provide a preliminary analysis of these two phenomena.

What explains the assassin's power to kill the victim? Although she has efficaciously *exercised* her power to effect the killing, *ex ante* she *has* the power to do so regardless of whether she chooses actually to exercise it. The assassin's possession of power could be explained by two sets of facts. First, if she were to favor the victim being killed, then she would consequently act, and the victim would be killed. Second, if she were to act on such a favoring attitude, the victim would die, whereas if she were not to so act, the victim would live. Hence, *ex ante*, a set of actions available to her is, within the given social-structural context, causally both necessary and sufficient for the outcome; and *ex post*, she has efficaciously exercised power insofar as her actions have caused the death. Insofar as the assassin favors the outcome, the welcome test is met; insofar as her action is necessary and sufficient for and hence causes the outcome, the intentional action test is met; and insofar as she intends the outcome, the link-age test is satisfied.

Yet such an explanation cannot cover all cases; in particular, it does not cover cases of causal overdetermination. Neither the necessity nor the sufficiency of an agent's actions is required. Consider the following.

Three Small Assassins: A victim is tied up in a car's trunk. There are three assassins, none of whom is strong enough to push the car over the cliff by himself, but any two together are sufficient. All three push, intentionally killing the victim.

No single assassin's action is either necessary or sufficient for the killing: if he were to have not pushed, the victim would still have been killed by the other two; if he had pushed on his own, the victim would not have been killed at all. Yet each plays a causal role in the killing. How is this exercise of agential causal power to be explained?

There exist several related approaches to explaining the causal role of individual conditions in cases of overdetermination.²² For our purposes, we do not need to decide between them; we need simply to note that causation can come in degrees and that in causally overdetermined cases, some causal conditions are merely partial causes. A simple approach for explaining this (for cases without preemption) is to analyze a condition's causal role in terms of a NESS test: a condition plays a causal role in case it (is not preempted and) is a *necessary element of a sufficient set* of conditions for the outcome, which set is a subset of the actual set of conditions on that occasion.²³ Here, the actual set of conditions comprises the pushing by the first, second, and third assassins. This set of three action events has three proper subsets sufficient for the outcome: the actions of the first and second, of the first and third, and of the second and third assassins. If the first assassin's action were sufficient for the outcome, it would be the *sole* necessary element of at least one sufficient subset of the actual set of conditions. And if it were necessary for the outcome, it would be an element of *all* sufficient subsets; as such, it would be fully causally efficacious. The fact that his action is insufficient does not imply he has no power; it implies he does not have strictly *unilateral* power to kill the victim—whatever power he has is a “power-with” others.²⁴ And the fact that his action is unnecessary does not imply he played no causal role: it implies his action was not decisive or a “full” cause. But since his action is a necessary element of two and only two of the three sufficient proper subsets, it is a *partially* efficacious cause.²⁵ *Ex ante*, the first assassin has some degree of power-with to effect the victim's killing.

Recognizing partial efficacy exposes the inadequacy of the widespread view that agents efficaciously exercise power only if they are *decisive*, i.e., only if, but for their action, the outcome would not occur. This view is presupposed by all who follow Robert Dahl in claiming that an agent exercises power over another only if she causes him to do something he “would not otherwise do” but for that exercise.²⁶ This view ignores the power one may exercise *with* others to effect an outcome—including power exercised *over* someone, causing them to alter

22 McDermott, “Redundant Causation”; Ramachandran, “A Counterfactual Analysis of Causation”; Hitchcock, “The Intransitivity of Causation Revealed in Equations and Graphs”; Schaffer, “Overdetermining Causes”; Halpern and Pearl, “Causes and Explanations: Part I” and “Causes and Explanations: Part II”; Braham, “Social Power and Social Causation”; Braham and van Hees, “Degrees of Causation”; and Pearl, *Causality*.

23 Wright, “Causation in Tort Law” and “Causation, Responsibility, Risk.”

24 Allen, “Rethinking Power”; and Abizadeh, “The Grammar of Social Power.”

25 Braham and van Hees, “Degrees of Causation.”

26 Dahl, “The Concept of Power,” 202–3. See Forst, *Normativity and Power*, 40.

their behavior—even when, as in overdetermined cases, one could not have unilaterally scuttled the outcome.²⁷

We now have a preliminary analysis of the most intuitively straightforward type of power: agential causal power. *Power*, in that the outcome is (or depends on being) favored by the agent. *Agential*, in that the outcome obtains (or would obtain) by way of the agent's *intentional actions*, where the agent's intention-in-action is appropriately *linked* to, because explained by, her favoring attitude. And *causal*, in that the agent effects (or would effect) the outcome via actions helping to cause it.

3. AGENTIAL NONCAUSAL POWER

The cases considered so far are cases of agential *causal* power. But casual efficacy—whether full or partial—is not only insufficient; it is also not necessary for effecting outcomes. It is not necessary because one can help ensure or see to an outcome without actually causing it.²⁸ Consider the following cases.

Preempted Poisoner: Assassin A injects a victim with a fatal poison. But before it takes effect, Assassin B shoots the victim, intentionally killing him instantly.

Preempted Shooter: A victim is sleeping. Many people want him dead before sunrise, but no one can tell whether he is sleeping or dead. To ensure he is dead, Assassin A will shoot the victim at 3 AM no matter what; her shot will be sufficient for the kill. But at 2:30 AM, Assassin B shoots the victim, intentionally killing him. At 3 AM, Assassin A shoots the victim, who, unbeknownst to Assassin A, is already dead.

In neither case is Assassin A's action necessary: even if she were not to poison or shoot, the victim would be killed. Nor indeed is Assassin B's action necessary in either case. However, Assassin A's action, like Assassin B's, *does* pass the NESS test: because her action is sufficient for the killing, it is the sole necessary element of a sufficient subset of the actual set of conditions. Yet in neither case does Assassin A's action *cause* the outcome: it operates too late to do so. These examples—of what David Lewis calls *late preemption*—show that the NESS test is insufficient for demonstrating causal efficacy because it is satisfied by not just preempting causes but also preempted potential causes.²⁹ A *complete causal test* must therefore add a further set of conditions ruling out preempted potential

27 Abizadeh, "The Power of Numbers."

28 Morriss, *Power*, 30–31.

29 Lewis, *Philosophical Papers*, 200–7.

causes. I do not defend a particular approach to fleshing out these further conditions; readers should supply whichever approach they deem most successful.³⁰

Ex ante, the preempted assassin has the power to see to it that the victim is killed; insofar as she exercises this power, she ensures the victim is killed. That the preempted assassin's action ensures the outcome is precisely why the preempting assassin's action is not necessary either on this occasion. The preempted assassin's action *effects* the victim's killing insofar as it satisfies the NESS test. But insofar as it fails the complete causal test (which rules out preempted potential causes), her action effects the killing without *causing* it.

A similar analysis can be provided of some types of invigilation.³¹ Consider the following case.

Invigilating Rainmaker: A farmer has a rainmaking machine. The machine cannot prevent rain, but on otherwise rainless days, it can be used to make it rain. On days in which it would rain naturally anyway, the natural causal process leading to rain renders the machine causally inert even if used. The farmer needs rain today and intends to use the machine if necessary. But it rains naturally today, so she leaves the machine idle.³²

Ex ante, the invigilating rainmaker has the power to ensure or see to it that it rains today. By acting on the conditional intention to use the rainmaking machine if necessary and refrain from using it if not, she exercises this power when she acts with that intention. It is true that the rainmaking machine gives her the power to cause it to rain in this general context. But today, on this particular occasion, given that nature causes it to rain, she does not have the power to cause it to rain. (If she were to try, she would fail, preempted by nature.) Therefore, *ex post*, the rainmaker effects the outcome on this occasion without causing it.

It might be objected that the invigilator cannot be said to effect the outcome *ex post* at all—on the grounds that, unlike the preempted poisoner and shooter above, she does not undertake any action. But invigilation is relevantly analogous to cases of late preemption. The difference is that the invigilator undertakes an intentional action whose intention has a *conditional* form. The point is perhaps clearer in the following case.

Invigilating Assassin: Two rival assassins want the same victim dead. Assassin A knows that Assassin B intends to kill the victim by 3 AM but

30 See Lewis, "Causation" and *Philosophical Papers*; McDermott, "Redundant Causation"; Ramachandran, "A Counterfactual Analysis of Causation"; Hitchcock, "The Intransitivity of Causation Revealed in Equations and Graphs"; and Pearl, *Causality*.

31 Pettit, "Freedom and Probability" and "Republican Freedom."

32 See Goldman, "Toward a Social Theory of Power"; and discussion in Morris, *Power*.

has doubts about his rival's efficacy so he secretly invigilates the killing, adopting the following conditional intention: if his rival succeeds by 3 AM, he will hold his fire, but if not, he will shoot the victim himself. Assassin *B* kills the victim before 3 AM. The invigilating assassin carries out his conditional intention by holding fire.

Forbearing from shooting is an intentional action that instantiates the invigilating assassin's conditional intention. The assassin has the power to ensure and hence *effect* the victim's killing and does ensure it by acting conditionally as he does, but given his rival's actions, he does not *cause* the outcome *ex post* on this occasion.

The similarity with late preemption stems from the fact that these invigilation cases are also ones of preempted causation: cases in which a potential causal process is preempted by another causal process. The difference is that in invigilation cases, the potential causal process is preempted not by preventing the preempted potential cause from causing the effect but by preempting the potential cause itself; they therefore instantiate not late but so-called *early* preemption.³³

It might be objected that neither late nor early preempted agents can ensure that *they themselves* kill their victim or make it rain and therefore cannot be said efficaciously to effect the relevant outcome. The premise is correct, but the conclusion does not follow. If the agent welcomes events only if they are caused by her, then the relevant outcome is an event-caused-by-her. True, the preempted shooter or invigilating rainmaker does not effect that outcome: she cannot ensure that she herself kills the victim or makes it rain. But if, as stipulated, her intention-in-action is not agent relative, she does effect the favored outcome: she ensures death or rain.

4. ELICITORY CAUSAL POWER

It is of course possible to cause outcomes without undertaking any intentional actions at all. Consider the following case.

Mafia Godfather: A mafia godfather's henchmen understand his overall objectives and therefore accurately anticipate his particular wishes. With nary a word from him, they anticipate and carry out what they believe to be his wishes. Given his objectives, were the mafioso cognizant of the threat posed by a police officer building an actionable case against him, he would want the officer assassinated. His henchmen, without so much

33 Lewis, *Philosophical Papers*, 200–7.

as apprising him of the situation, assassinate the officer. Whatever the godfather were to want, his henchmen would seek to realize.

The mafioso's objectives are fulfilled, but not by way of his intentional actions. If exercising power implies doing so by way of intentional actions, then, because the agent-outcome relation fails the intentional action test, the mafioso cannot be said to be exercising agential power here. Yet I take it that, recognizably, the mafioso's objectives are nevertheless fulfilled in virtue and by way of his power, and not as a mere side effect unrelated to his preferences or objectives: the outcomes are not only welcome to him (satisfying the *ex post* welcome test); they are caused by and counterfactually vary with his preferences (satisfying the *ex ante* welcome test). Indeed, the mafioso may very well prefer to wield power passively in this way—not only to save himself the effort but also because it maintains plausible deniability!

What explains the mafioso's power to obtain his preferences without acting? Let us begin by filling out some implicit background details. There is, firstly, a hierarchical social relation between the godfather and each henchman: they all recognize him as boss, with the power to issue orders to them, etc. This relation, secondly, is also structural: a set of "regulators"—not just he and any given particular henchman but also the other henchmen, others in the organization, indeed other crime groups and the police—also recognize his position over the henchman, and their recognition and treatment of the former as the boss, and the common knowledge that others will do so as well, stabilize and reinforce the relation. So the mafioso clearly has *agential* power in virtue of his structural position: he can order his henchmen to do things. And insofar as his power to obtain outcomes—for example by ordering them around—is due to the position he occupies over his henchmen, his power is "structural" in the broad sense I set aside.

Our question is how his position translates into elicitory power. We can consider several variants of the Mafia Godfather case, each with its own explanation. On the first, which we can call the Vengeful Mafioso variant, what explains his elicitory power is his henchmen's anticipation of his reaction should they not fulfill his objectives: if they were to not assassinate the officer, the mafioso may kill them for incompetence or disloyalty. This is what Carl Friedrich calls the "rule of anticipated reactions," referring to the reactions of the *power holder* himself.³⁴ The Friedrichian interpretation may lead some to question whether the Mafia Godfather case supports the notion of elicitory power. If what instills fearful anticipation and loyalty is the mafioso's current and known disposition to exact vengeance, then it might be objected that he

34 Friedrich, *Constitutional Government and Politics*, 16–18, and *Man and Government*, 199–215.

does exercise agential power via his intentional actions after all. There are three versions of this objection.

First, the reason why the vengeful mafioso's henchmen anticipate his wishes and vengeance might after all be due to the mafioso's *past* intentional actions and ensuing reputation.³⁵ However, to have intentionally acted in the past to create a reputation in virtue of which one now passively elicits outcomes is not *now* actively to exercise agential power. It is to have exercised agential power in the past to effect a structure in which one now passively elicits outcomes. Others' knowledge of one's disposition for unwanted reactions, moreover, is not always caused by one's past actions. Sometimes it is known because one is a token of a certain type, for example, by occupying a certain social-structural position. Consider the following case.

Capitalist Giant: A capitalist owns and controls a giant corporation. His aim is to maximize profits; to fulfill this aim, he must invest in the jurisdiction with the lowest corporate taxes. If the corporation's current jurisdiction raises corporate taxes, he will move the corporation elsewhere, with catastrophic economic consequences for his current jurisdiction. The capitalist would not intend to inflict these harms; he would merely intend to maximize profits. The incentive structure of capitalist corporations is well known to the government; to avoid catastrophe, it invariably sets corporate taxes at a low rate.³⁶

The capitalist giant elicits outcomes in virtue of the government's anticipation of his hypothetical future actions should it act contrary to his preferences. However, the government knows his preferences and consequently his disposition, not because of his past actions but because of the capitalist's type and position in the economic structure.

A second version of the objection is that the mafioso's disposition to avenge failures reflects a conditional standing intention to do so, in which case he *does* undertake a negative intentional action: he is an invigilator. I concede that *if* so, then the case of anticipated reactions would be one in which the mafioso exercises agential power. But although some mafioso might have such a standing intention, another might not: he might have a disposition without a standing intention, for example, if had not yet decided or adequately considered what to do if his henchmen betrayed his trust. The vengeful mafioso I have in mind is of the latter type.

35 Dowding, "Resources, Power and Systematic Luck," 316.

36 See Lindblom, "The Market as Prison"; and Barry, "Capitalists Rule Ok?"

Third, some might worry that the vengeful mafioso is poor evidence for elicitory power because whatever power he has seems parasitic on his *agential* power to exact revenge. Although he does not fulfill his preferences by way of his *actual* intentional actions, he does seem to realize them by way of *hypothetical* intentional actions—which he would undertake if he were betrayed. And it might be argued that the category of agential power should be expanded to include wielding power by way of hypothetical intentional actions of this kind, without invoking a separate category of elicitory power.³⁷ The problem with this suggestion is twofold. First, the vengeful mafioso would wield the same power even if he did not have or was not disposed to exercise the agential power to exact revenge, just as long his henchmen believed that he has and is disposed to exercise such power. Second, the objection fails to account for the other variants of the Mafia Godfather case, to which we now turn.

Sometimes agents have elicitory power in virtue of the anticipated reactions of other agents, not their own reactions. Consider the case of the Figurehead Mafioso. The figurehead has no capacity to exact vengeance on his henchmen, and they know this. Their loyalty is explained rather by the larger social context of rivalry between mafia gangs: if the figurehead mafioso were to fall, his henchmen would fall with him. What motivates them is the anticipated actions of third-party “peripheral actors” or “regulators,” not the godfather’s. The henchmen act on the basis of what they take to be his preferences, not necessarily because they deem his judgment about such matters superior to their own but because they use him as a focal-point coordinating mechanism: their gang’s standing requires an efficient way to coordinate their actions without acting at cross-purposes.³⁸ This mafioso elicits outcomes in virtue of his social-structural position in the network of mafia hierarchies. He does not have the agential power to effect the outcomes nor even effectively to retaliate against failures, but nevertheless elicits those outcomes by his structural power.

Some cases of elicitory power do not depend on the anticipation of *anyone’s* future actions. Consider the case of the Old Mafioso, whose gang has successfully wiped out all rivals and is at little risk of failing. He is so old and decrepit now that he cannot leave his bed and no longer has the power to extract deadly vengeance against his henchmen. But his henchmen are so used to anticipating and carrying out his wishes—except extracting vengeance against fellow henchmen—they can no longer imagine doing otherwise: they cannot fathom their life’s purpose or meaning without serving their godfather. Serving has become second nature. The old mafioso no longer has and hence could not

37 I thank David Estlund for this suggestion. See Barry, “Capitalists Rule Ok?” 178–80.

38 Schelling, *The Strategy of Conflict*.

even hypothetically exercise the agential power to avenge the outcomes, but he still has the passive power to elicit them: outcomes vary with his preferences because his preferences help cause them.

Similarly, consider the following case.

Charismatic Prophet: A prophet has such a charismatic presence that many fall under her spell merely from being in her presence. She so inspires her admirers that they anticipate and seek to fulfill her every wish even before she herself has become conscious of them, not because they seek future reward or approval but because they take joy in serving her.

The charismatic prophet elicits outcomes thanks to her elicitory causal power: some features of her besides her intentional actions—such as her physiognomy, odor, and unconscious demeanor—cause others to anticipate her preferences and fulfill them. In a social context, such power also at least partly exists in virtue of the agent's position within a social structure—for example, the structure of norms and mental schemas according to which a woman's physiognomy, odor, and demeanor tend to be interpreted. But her preferences' causal role in eliciting outcomes is not due to anticipated reactions.

Here is another example.³⁹ Imagine we live in a society in which, thanks to the racial structure, people are very attentive and sympathetic to how light-skinned people wish to be treated: most are disposed to detect and satisfy light-skinned people's preferences for how to be treated. You already have light skin, so this is quite pleasant for you. But I have dark skin and, predictably, have not been treated as I had wished. However, my skin is not too dark, I have money, and I can and do buy effective skin-lightening creams. My prior intention and my intention-in-action in buying and applying these creams are to get people to treat me as I wish, which is exactly what happens. So I have efficaciously exercised my agential power to effect outcomes satisfying my preferences for how others treat me. Yet what I have done is merely, by way of my intentional actions, to try to mimic a power you already have and wield but not by way of any intentional actions. You fulfill the same type of preferences for yourself without having to do anything. I have to expend energy exercising agential power just to mimic your elicitory, structural power.

In sum, when agents' preferences help cause outcomes independently of their intentional actions, they elicit those outcomes. They wield elicitory power.

39 I thank Lucas Stanczyk for suggesting an example like this.

5. ELICITORY NONCAUSAL POWER

The most controversial type of power I defend here is perhaps elicitory non-causal power. But once we acknowledge that one can effect outcomes by exercising agential *noncausal* power, on the one hand, and elicit outcomes via *elicitory* causal power, on the other, then there is no reason to deny the combination of elicitory and noncausal power. Not only can social power take this form; it is one of the most important kinds of social power.

How can agents passively wield power to elicit outcomes by way of neither intentional actions nor any causal role? One obvious way (analogous to non-causal agential power) is if their preferences help to ensure the outcome but are preempted. Consider the case of the Preempted Mafioso, whose preferences are not even the outcome's *actual* cause. Imagine the henchmen sometimes cannot discern his preferences as easily and swiftly as they can his son's. Since the son shares his father's objectives and invariably wants the same, the henchmen treat (their sense of) what the son wants as a kind of oracle for his father's wishes. Now imagine they think the decision whether to assassinate is urgent but that trying to discern the father's preference risks too much delay; so they act instead on the basis of discerning the son's preference. Here it is the son's preference, not the father's, that causes the outcome, but had the son not preferred it, the henchmen would have later discerned the father's preference and carried out the assassination anyway. The son's preference preempted the father's potential causal role, but the preempted mafioso nevertheless has (passively) elicited the outcome, in a manner similar to how invigilators (actively) effect outcomes: his preference passively invigilated the outcome, so to speak, by ensuring it would happen.

Sometimes, moreover, agents have the power to fulfill their preferences thanks to the causal role of others' actions or preferences rather than their own, even without their own preferences being preempted causes. Consider the following case.

Immobile Little Capitalist: A capitalist owns and controls a small business. Because his aim is to maximize profits, his incentive structure normally propels him to invest in the jurisdiction with the lowest corporate tax rates. But due to family responsibilities, he is now immobile and will stay in his current jurisdiction whatever the tax rate. There are, however, hundreds of other little capitalists in his jurisdiction who are mobile: if the government raises corporate taxes, they will all move their businesses elsewhere, with catastrophic economic consequences for the current jurisdiction. The incentive structure of capitalists is well known to the government; to avoid catastrophe, it sets corporate taxes at a low rate.

The outcome is welcome to the little capitalist, but his preferences (and anticipated reactions) play no causal role in effecting it: the outcome is caused by the preferences and anticipated reactions of other, mobile little capitalists. Nothing about the immobile little capitalist—not even the position he occupies in the social structure as a capitalist business owner—helps cause the outcome. But in virtue of the position he occupies, the immobile little capitalist obtains the tax outcomes he wishes. He has the power to fulfill his preferences thanks to his position and what *others* systematically prefer and would do, and thanks to the systematic correlation of their preferences with his—but independently of his intentional actions or any causal role (or even preempted causal role) for his preferences.

Why should this count as having and wielding structural power rather than merely as good luck and benefitting? Because the relevant outcomes systematically vary with the agent's preferences. Imagine the economic situation changes, such that the profits of little capitalists come to depend on a highly educated, highly skilled workforce. And imagine that producing such a workforce requires public state investment in early education and training that can be funded only via higher corporate taxes.⁴⁰ Now the little capitalists all prefer jurisdictions with significant educational investment funded by higher taxes. To prevent the little *mobile* capitalists from leaving, the state raises corporate taxes, just as the little *mobile and immobile* capitalists have come to prefer. This is no mere benefitting. It is benefitting that counterfactually varies with the *mobile and immobile* capitalists' preferences—even though the preferences of only the former actually *cause* the varied benefits. (The preferences of the latter, in turn, vary with those of the former because of their similar position in the economic structure.) The difference between merely benefitting versus wielding elicitory, structural power turns on this key point.

Therefore, not all beneficial outcomes may be imputed to an agent's elicitory power. Consider a neighbor who just happens to splash water on the plants in your parched garden. That the outcome is beneficial and welcome to you is not sufficient for showing you have elicited it in virtue of your power. We cannot discern elicitory power just by observing that the outcome is *ex post* welcome on a particular occasion: what matters is whether such outcomes would vary with hypothetical variations in your preferences. For there to be such variation in principle—for the splashing to count as elicited by your power—then at least one of two conditions must hold. Either your preferences were a full, partial, or preempted cause of the splashing; for example, your neighbor knew you wanted your plants watered and was therefore moved to splash water because,

40 I thank Jacob Levy for suggesting this example.

due to your standing in the community, she wants to curry your favor. Or, if your preferences did not cause or help ensure the splashing, then, at least on some occasions in which the splashing might have happened, if you had counterfactually preferred that it not happen, then it would not. Otherwise, you have merely benefitted. The welcome watering is elicited by your structural power only if your neighbor's actions somehow counterfactually track your preferences thanks to your position in a social structure (such as the property regime).

More generally, whereas agential power requires satisfying the welcome, intentional action, and linkage tests, elicitory power requires satisfying only the welcome test. But to have efficaciously wielded elicitory power *ex post*, it is necessary *but not sufficient* to satisfy the *ex post* welcome test. The possession and hence efficacious wielding of elicitory power requires also satisfying the *ex ante* welcome test: not only must the actual outcome be welcome *ex post* on this particular occasion; outcomes must also systematically vary counterfactually with the agent's preferences in this context in general. This is what explains the difference between efficaciously wielding structural power and mere benefitting.

It also explains the difference from *systematic* mere benefitting. Consider an incumbent infrastructural regime such as an existing clean water system whose continuing functioning requires no maintenance during my lifespan.⁴¹ I "systematically" benefit from it in the sense that my actual preferences for clean water are continually and habitually fulfilled. But I do not thereby have elicitory or structural power to fulfill my preference for clean water: the provision of the benefit does not counterfactually vary with my preferences. Were I counterfactually to prefer clean kombucha rather than water, the infrastructure would still furnish water.

6. CONCLUSION

Why is it important to recognize nondecisive, noncausal, and elicitory and structural categories of power? Peter Morriss argues that the concept of power serves three purposes: the practical purpose of determining how to fulfill our aims; the moral purpose of allocating responsibility and blame; and the evaluative purpose of appraising social arrangements.⁴² Each of these contexts illuminates the significance of recognizing the types of power defended here.

Begin with the practical purpose of trying to identify the agents over whom it would be useful to wield power. If someone has *agential* power over issues

41 I thank Matthew Noah Smith for the example.

42 Morriss, *Power*, 37–42.

that concern you, then of course it makes perfect instrumental sense to wield power over them—to get them to do something they otherwise they might not do, to reprise Dahl's classic formulation. But it *also* makes good instrumental sense to wield power over agents with elicitory, structural power: shaping their preferences would be another way to obtain or ensure your preferred outcomes—albeit without getting them to *do* anything.

Next consider moral theory. It is a platitude that power and responsibility go hand in hand. It is also widely held that to be normatively responsible and appropriately blamed, two conditions must be met: first, one must be normatively *competent*; and second, that for which one is responsible must be under one's *control*. A plausible account of normative competence equates it with *rational agency*: the capacity to grasp concepts and recognize facts as reasons; to reflectively assess what reasons one has; and to respond, in one's intentional states and actions, to those reflectively assessed reasons.⁴³

The widespread restriction of power to agential power has often been motivated by a particular gloss on the second, control condition: that one can bear normative responsibility and appropriately be blamed only for what is under one's *voluntary* control. On this view, one can be (derivatively) responsible only for the consequences of one's choices and intentional actions.⁴⁴ Thus, if power is linked to responsibility and serves to identify those responsible, then it seems the notion of power at stake must be agential—operating by way of intentional actions.

This “voluntarist” theory of responsibility and blame, however, is highly controversial at best; indeed, I believe it to be mistaken, but I limit myself to showing why we should not burden a theory of power with it.⁴⁵ The problem is that it sits poorly with our actual responsibility and blaming practices. There is nothing more pedestrian than resenting or feeling indignation towards those who take pleasure in or desire others' pain, bear evil hopes or desires for others, fail to take others' well-being into account in deliberation, or believe that members of our social group do not merit respect or are less capable than we are.⁴⁶ And there is no denying that people resent friends for regularly forgetting a special occasion, failing to notice ethically salient features of the current circumstances, or involuntarily betraying indifference or contempt. Yet unlike

43 Wallace, *Responsibility and the Moral Sentiments*, 157–58; and Skorupski, *The Domain of Reasons*, 21–23, 59.

44 Wolf, *Freedom Within Reason*; and Wallace, *Responsibility and the Moral Sentiments*.

45 I leave aside incompatibilist accounts equating control with metaphysical *free will*, which, combined with the supposition of causal determinism, would evacuate any talk of normative responsibility and blame. See Fischer and Ravizza, *Responsibility and Control*.

46 Adams, “Involuntary Sins.”

bodily movements, we typically cannot directly will or choose to believe, feel, or desire: many thoughts, emotions, and desires are spontaneous, involuntary responses to the world. In short, we often blame people and hold them responsible for involuntary states (and behaviors). True, sometimes our reactive blaming responses target agents' past, character-shaping choices, but often we resent the intentional state itself, not its genetic history.⁴⁷

A plausible explanation for these practices is that we appropriately hold agents responsible for intentional states insofar as these are rationally responsive to agents' own reflective evaluations or judgments. On this construal of the control condition—which fits much more neatly with the first, rational agency condition—we can be blameworthy even for involuntary states or behaviors insofar (and only insofar) as they are judgment sensitive.⁴⁸ What “controls” or guides our states is not our will but our evaluative judgments or the mechanisms constituting our rational agency.⁴⁹

The upshot is that there is no responsibility-related reason for restricting the power concept to agential power. Indeed, there is every reason for it to encompass elicitory power: the preferences on the basis of which the welcome test analyzes elicitory power are judgment-sensitive evaluative states.⁵⁰ In other words, one normative payoff of the theory defended here is that it opens the door for power structure analyses that more closely track actual practices and better align with more plausible theories of responsibility and blame.

Another payoff is to provide a plausible framework for thinking about how power may be related to the different forms of holding people responsible. Here I can merely sketch out this possibility. Consider the blaming emotions that, if we follow the Strawsonian tradition, are constitutive of blame and holding people morally responsible. The content of anger, disdain, or shame, on the one hand, and resentment, indignation, or guilt, on the other, is that its target has violated an expectation, standard, or requirement. But unlike the first trio, the second trio specifically concerns requirements *owed* to others; as such, they are “reactive” or vindictory, second-personal emotions in the sense that they inherently

47 Smith, “Responsibility for Attitudes.”

48 Smith, “Responsibility for Attitudes.”

49 Fischer and Ravizza, *Responsibility and Control*.

50 Hausman, *Preference, Value, Choice, and Welfare*. True, some sources of preferences may be judgment insensitive. But because the preferences in question are total or *all-things-considered* evaluative rankings incorporating all relevant sources, they remain judgment sensitive. Even if my preference for obtaining vanilla over chocolate sorbet were currently based solely on brute taste, if I were now to become aware of a normative reason against obtaining vanilla, I may, all things considered, come to prefer obtaining chocolate (despite the partial preference for vanilla on taste grounds).

demand a response from their target—whether justification, excuse, apology, compensation, or measures to prevent future violations.⁵¹ Expressly to hold someone morally accountable for a perceived blameworthy violation is, in the first instance, to demand such responses by way of expressing reactive blaming emotions (or the belief that they are appropriate) and/or to impose sanctions.

There is no reason to attribute blanket immunity against these responses to agents whose power falls short of decisive, causal agential power. Nevertheless, we can now see how, depending on the type of power, some modes of holding responsible may be inappropriate. Due to collective action problems, for example, agents with nondecisive power may be unable to implement prevention measures and hence not be liable. Their nondecisiveness may also furnish an excuse for violations, alongside standard excuses like ignorance. Agents with causal structural power may, in virtue of their causally efficacious preferences, be open to all the accountability demands and even to interpersonal or social sanction but, because choice is missing, not be open to coercively imposed political sanction. And whether those with noncausal (because preempted) power—agential or structural—are deemed immune from demands for compensation may depend on which views about circumstantial moral luck are justified. Moreover, I suspect that one whose noncausal structural power stems solely from sharing preferences with others is not the appropriate target of reactive blame at all—but may very well be the appropriate target of some of the *demands* associated with them (such as prevention measures) and perhaps even nonreactive forms of blame.⁵²

Finally, from an evaluative perspective, why is it important to recognize structural power, including noncausal structural power, as a category of power—rather than as merely good luck or privilege? True, sometimes “privilege” amounts to mere systematic benefitting, not power. But sometimes privilege amounts to more, and when it does, we should recognize it as structural *power* and not merely “privilege” because structural power is a wider category. It is not inherently hierarchical, inequalitarian, or evil: it can be widely dispersed, equally shared, reciprocal, and in many cases of normative value. Consider a society governed by an effective norm that others offer their seat to any person facing physical hardship; a Good Samaritan norm that others nearby give aid, shelter, or protection to any imperiled person; or laws providing for basic health care to any person in need. All persons in such a society would have, in virtue of their position in the corresponding social structures, the structural power to have these fundamental needs met—independently of their own intentional actions.

51 Strawson, *Freedom and Resentment and Other Essays*; Wallace, *Responsibility and the Moral Sentiments*; and Darwall, *Second-Person Standpoint*.

52 Watson, “Two Faces of Responsibility.”

Furthermore, many egalitarians prize political equality, frequently understood to demand equal power over binding political decisions, while neorepublicans seek to minimize relations of domination, characterized as arbitrary relations of power-over.⁵³ But if the theory defended here stands to reason, then political equality and nondomination may require attending to inequalities or arbitrariness in noncausal and structural forms of power as well. To miss these phenomena as instances of power is to obscure one of the central concerns of social theory and normative political philosophy: the distribution of power.⁵⁴

McGill University
arash.abizadeh@mcgill.ca

- 53 For discussion of the former, see Beitz, *Political Equality*; Wilson, *Democratic Equality*; and Wodak, "What Is the Point of Political Equality?" For the latter, see Pettit, *A Theory of Freedom*; and Lovett, *A General Theory of Domination and Justice*.
- 54 Funding was provided by Social Sciences and Humanities Research Council (SSHRC) of Canada grant 435-2025-0832 and Fonds de recherche du Québec – Société et culture (FRQSC) grant 2023-SE3-318011. For comments on previous drafts, I am grateful to Arthur Applbaum, Federica Berdini, Yuna Blajer de la Garza, Adrian Blau, Matthew Braham, Garrett Cullity, Jorah Dannenberg, Robin Douglass, Keith Dowding, Yiftah Elazar, David Enoch, David Estlund, Mollie Gerver, Pablo Gilabert, Valéry Giroux, Carole Gould, Alex Gourevitch, Amanda Greene, Allen Habib, Alon Harel, Sally Haslanger, Sean Ingham, Dongxian Jiang, David Johnston, Steven Klein, Kyuree Kim, Yunhyae Kim, Niko Kolodny, Thomas Krödel, James LaBelle, Holly Lawford-Smith, Chad Lee-Stronach, Jeffrey Lenowitz, Jacob Levy, Éliot Litalien, Catherine Lu, Karuna Mantena, Mara Marin, Shal Marriott, Jonathan Leader Maynard, Victor Muñoz-Fraticelli, Josh Ober, Armando Jose Perez-Gea, Ryan Pevnick, Veronica Ponce, Samantha Puzzi, Will Roberts, Andrea Sangiovanni, Shlomi Segal, Joshua Simon, Matthew Noah Smith, Jules Solomone-Sehr, Nic Southwood, Lucas Stanczyk, Jordan Walters, Leif Wenar, Daniel Weinstock, Vanessa Wills, Yves Winter, Yi Yang, several anonymous referees, and participants at the McGill Research Group on Constitutional Studies (RGCS) workshop on February 18, 2021; the Stanford Political Theory Workshop on February 19, 2021; the Columbia Political Theory Workshop on March 31, 2021; the Centre de recherche en éthique (CRÉ) lunch series at Université de Montréal on September 21, 2021; the Kings College London Department of Political Economy Seminar on September 29, 2021; the Jerusalem Seminar in Moral and Political Philosophy on April 24, 2022; the RGCS/CRÉ/Groupe de recherche interuniversitaire en philosophie politique manuscript workshop on August 15–17, 2022; City University of New York (CUNY) Center for Global Ethics and Politics on October 13, 2022; the Annual Meeting of the Philosophy, Politics, and Economics Society in New Orleans on November 3–5, 2022; the Brown University Philosophy Department manuscript workshop on March 25, 2023; the Australian National University Conference in Honour of Robert Goodin on August 8–9, 2023; the University of Sydney Philosophy Seminar Series on August 14, 2023; the Hamburg University Collective Decision-Making Program Interdisciplinary Research Seminar on April 4, 2024; and the Hamburg University Collective Decision-Making Program manuscript workshop on April 8, 2024.

REFERENCES

- Abizadeh, Arash. "The Grammar of Social Power: Power-to, Power-with, Power-despite and Power-over." *Political Studies* 71, no. 1 (2023): 3–19.
- . "The Power of Numbers: On Agential Power-with-Others Without Power-over-Others." *Philosophy and Public Affairs* 49, no. 3 (2021): 290–318.
- Adams, Robert Merrihew. "Involuntary Sins." *Philosophical Review* 94, no. 1 (1985): 3–31.
- Allen, Amy. "Rethinking Power." *Hypatia* 13, no. 1 (1998): 21–40.
- Ball, Terence. "Models of Power: Past and Present." *Journal of the History of the Behavioral Sciences* 11, no. 3 (1975): 211–22.
- . "Power, Causation, and Explanation." *Polity* 8, no. 2 (1975): 189–214.
- Barry, Brian. "Capitalists Rule Ok? Some Puzzles About Power." *Politics, Philosophy, and Economics* 1, no. 2 (2002): 155–84.
- . "The Uses of 'Power.'" *Government and Opposition* 23, no. 3 (1988): 340–53.
- Beitz, Charles R. *Political Equality: An Essay in Democratic Theory*. Princeton University Press, 1989.
- Bourdieu, Pierre. *La Domination Masculine*. Seuil, 1998.
- Braham, Matthew. "Social Power and Social Causation: Towards a Formal Synthesis." In *Power, Freedom, and Voting: Essays in Honour of Manfred J. Holler*, edited by Matthew Braham and Frank Steffen. Springer, 2008.
- Braham, Matthew, and Martin van Hees. "Degrees of Causation." *Erkenntnis* 71, no. 3 (2009): 323–44.
- Dahl, Robert A. "The Concept of Power." *Behavioral Science* 2, no. 3 (1957): 201–15.
- . "Power." In *International Encyclopedia of the Social Sciences*, edited by David L. Sills. Macmillan, 1968.
- Darwall, Stephen. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Harvard University Press, 2009.
- Debnam, Geoffrey. "Nondecisions and Power: The Two Faces of Bachrach and Baratz." *American Political Science Review* 69, no. 3 (1975): 889–99.
- Dowding, Keith. *Rational Choice and Political Power*. Edward Elgar, 1991.
- . "Resources, Power and Systematic Luck: A Response to Barry." *Politics, Philosophy, and Economics* 2, no. 3 (2003): 305–22.
- Dretske, Fred. *Explaining Behavior: Reasons in a World of Causes*. Massachusetts Institute of Technology Press, 1988.
- Elder-Vass, Dave. *The Causal Power of Social Structures: Emergence, Structure and Agency*. Cambridge University Press, 2010.
- Fischer, John Martin, and Mark Ravizza. *Responsibility and Control: A Theory*

- of *Moral Responsibility*. Cambridge University Press, 1998.
- Forst, Rainer. *Normativity and Power: Analyzing Social Orders of Justification*. Oxford University Press, 2017.
- Friedrich, Carl J. *Constitutional Government and Politics*. Harper and Brothers, 1937.
- . *Man and Government*. McGraw-Hill, 1963.
- Gädeke, Dorothea. “Does a Mugger Dominate? Episodic Power and the Structural Dimension of Domination.” *Journal of Political Philosophy* 28, no. 2 (2020): 199–221.
- Goldman, Alvin I. “Toward a Social Theory of Power.” *Philosophical Studies* 23, no. 4 (1972): 221–68.
- Halpern, Joseph Y., and Judea Pearl. “Causes and Explanations: A Structural-Model Approach. Part I: Causes.” *British Journal for the Philosophy of Science* 56, no. 4 (2005): 843–87.
- . “Causes and Explanations: A Structural-Model Approach. Part II: Explanations.” *British Journal for the Philosophy of Science* 56, no. 4 (2005): 889–911.
- Harris, Cheryl I. “Whiteness as Property.” *Harvard Law Review* 106, no. 8 (1993): 1707–91.
- Hasan, Rafeeq. “Republicanism and Structural Domination.” *Pacific Philosophical Quarterly* 102, no. 2 (2021): 292–319.
- Haslanger, Sally. “Oppressions: Racial and Other.” In *Racism in Mind*, edited by Michael P. Levine and Tamas Pataki. Cornell University Press, 2004.
- Hausman, Daniel M. *Preference, Value, Choice, and Welfare*. Cambridge University Press, 2012.
- Hayward, Clarissa Rile. *De-Facing Power*. Cambridge University Press, 2000.
- . “On Structural Power.” *Journal of Political Power* 11, no. 1 (2018): 56–67.
- Heil, John. *The Universe as We Find It*. Oxford University Press, 2012.
- Hitchcock, Christopher. “The Intransitivity of Causation Revealed in Equations and Graphs.” *Journal of Philosophy* 98, no. 6 (2001): 273–99.
- Ingthorsson, Rognvaldur D. “Causal Production as Interaction.” *Metaphysica* 3, no. 1 (2002): 87–119.
- Isaac, Jeffrey C. *Power and Marxist Theory: A Realist View*. Cornell University Press, 1987.
- Kuykendall, Davis White. “In Defense of the Agent and Patient Distinction: The Case from Molecular Biology and Chemistry.” *British Journal for the Philosophy of Science* 75, no. 2 (2024): 443–63.
- Laswell, Harold, and Abraham Kaplan. *Power and Society*. Yale University Press, 1950.
- Lewis, David. “Causation.” *Journal of Philosophy* 70, no. 17 (1973): 556–67.

- . *Philosophical Papers*, vol. 2. Oxford University Press, 1986.
- Lindblom, Charles E. "The Market as Prison." *Journal of Politics* 44, no. 2 (1982): 324–36.
- Lovett, Frank. *A General Theory of Domination and Justice*. Oxford University Press, 2010.
- March, James G. "An Introduction to the Theory and Measurement of Influence." *American Political Science Review* 49, no. 2 (1955): 431–51.
- Marmodoro, Anna. "Aristotelian Powers at Work: Reciprocity Without Symmetry in Causation." In *Causal Powers*, edited by Jonathan D. Jacobs. Oxford University Press, 2017.
- Marsh, David. "Interest Group Activity and Structural Power: Lindblom's *Politics and Markets*." *West European Politics* 6, no. 2 (1983): 3–13.
- McDermott, Michael. "Redundant Causation." *British Journal for the Philosophy of Science* 46, no. 4 (1995): 523–44.
- Mill, John Stuart. *The Subjection of Women*. Vol. 21 of *The Collected Works of John Stuart Mill*, edited by J. M. Robson. Toronto University Press, 1984.
- Morriss, Peter. *Power: A Philosophical Analysis*. 2nd ed. Manchester University Press, 2002.
- Nagel, Jack H. *The Descriptive Analysis of Power*. Yale University Press, 1975.
- O'Connor, Timothy. *Persons and Causes: The Metaphysics of Free Will*. Oxford University Press, 2002.
- Partridge, P. H. "Some Notes on the Concept of Power." *Political Studies* 11, no. 2 (1963): 107–25.
- Pearl, Judea. *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge University Press, 2009.
- Pettit, Philip. "Freedom and Probability: A Comment on Goodin and Jackson." *Philosophy and Public Affairs* 36, no. 2 (2008): 206–20.
- . "Republican Freedom: Three Axioms, Four Theorems." In *Republicanism and Political Theory*, edited by Cécile Laborde and John Maynor. Blackwell, 2008.
- . *A Theory of Freedom: From the Psychology to the Politics of Agency*. Oxford University Press, 2001.
- Ramachandran, Murali. "A Counterfactual Analysis of Causation." *Mind* 106, no. 422 (1997): 263–77.
- Roy, William G. *Socializing Capital: The Rise of the Large Industrial Corporation in America*. Princeton University Press, 1997.
- Russell, Bertrand. *Power: A New Social Analysis*. Routledge, 2004.
- Schaffer, Jonathan. "Overdetermining Causes." *Philosophical Studies* 114, no. 1 (2003): 23–45.
- Schelling, Thomas C. *The Strategy of Conflict*. 2nd ed. Harvard University Press,

1980.

Searle, John R. *Making the Social World: The Structure of Human Civilization*. Oxford University Press, 2010.

Simon, Herbert A. *Models of Man*. Wiley, 1957.

Skorupski, John. *The Domain of Reasons*. Oxford University Press, 2010.

Smith, Angela M. "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115, no. 2 (2005): 236–71.

Stone, Clarence N. "Systemic Power in Community Decision Making: A Restatement of Stratification Theory." *American Political Science Review* 74, no. 4 (1980): 978–90.

Strange, Susan. *States and Markets*. 2nd ed. Continuum, 1994.

Strawson, P. F. *Freedom and Resentment and Other Essays*. Routledge, 2008.

Vrousalis, Nicholas. "The Capitalist Cage: Structural Domination and Collective Agency in the Market." *Journal of Applied Philosophy* 38, no. 1 (2021): 40–54.

Wallace, R. Jay. *Responsibility and the Moral Sentiments*. Harvard University Press, 1996.

Ward, Hugh. "Structural Power: A Contradiction in Terms?" *Political Studies* 35, no. 4 (1987): 593–610.

Wartenberg, Thomas E. *The Forms of Power: From Domination to Transformation*. Temple University Press, 1990.

Watson, Gary. "Two Faces of Responsibility." *Philosophical Topics* 24, no. 2 (1996): 227–48.

White, D. M. "Power and Intention." *American Political Science Review* 65, no. 3 (1971): 749–59.

Wilson, James Lindley. *Democratic Equality*. Princeton University Press, 2019.

Wodak, Daniel. "What Is the Point of Political Equality?" *Philosophical Review* 133, no. 4 (2024): 367–413.

Wolf, Susan. *Freedom Within Reason*. Oxford University Press, 1990.

Wright, Richard W. "Causation in Tort Law." *California Law Review* 73 (1985): 1735–828.

———. "Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts." *Iowa Law Review* 73 (1988): 1001–77.

Wrong, Dennis H. *Power: Its Forms, Bases, and Uses*. Routledge, 2017.

BEING WRONGED AND UNDERSTANDING MORAL WRONGNESS

Daniel Vanello

THE AIM of this article is to articulate and defend the intuition that the experience of being morally wronged affords one a distinctive understanding of the moral wrongness of what one experiences. In section 1, I clarify and motivate this claim. In section 2, I argue that the relevant epistemic good is the understanding of what it is like to endure the emotional pain of someone who must make sense of one's autobiography and thus sense of self as a person who has been wronged in the relevant way. In section 3, I argue that the epistemic ability that is exercised in the generation of this understanding is the ability to reflect on one's personal experiences and articulate a generalizable understanding of the relevant moral wrongness. In section 4, I argue that this epistemic good and related epistemic ability allow one to make claims about the objective wrong-making features of the experienced event and is not confined to merely making claims about features of the experience. In section 5, I argue that this epistemic ability should not be conflated with the epistemic abilities of the sort that can be successfully exercised independently of having undergone the relevant experiences. This then allows me to spell out the relationship between the epistemic ability exercised by those who experience the relevant wrong and the various epistemic abilities that can be possessed by those who do not undergo the relevant experiences. I conclude with section 6, where I argue that the account of this article strengthens the view that those who do not undergo experiences of being wronged in the relevant way have the responsibility to seek out and acquire knowledge of these wrongs in order not to perpetuate them.

1. MOTIVATING THE CLAIM

Consider the following statements made by feminist philosophers in the standpoint epistemological tradition. Uma Narayan writes, "A very important component of what constitutes the epistemic privilege of the oppressed has to do with knowledge that is at least partly constituted by and conferred by the emotional responses of the oppressed to their oppression." Elizabeth Anderson writes, "The

epistemic privilege of the oppressed resides in their privileged access to certain experiences, which give them information especially revelatory of fundamental truths about society." Alison Wylie writes that "the experience of exclusion or marginalization may itself be a source of insight" into what oppression is. Manon Garcia relies on the work of Simone de Beauvoir to make an analogous claim about submission. The reason she relies on de Beauvoir's work is that de Beauvoir, in virtue of "her privileged social position," "has access to the experience of submission without being silenced like many submissive women."¹

How should we understand these statements? Recent philosophical engagements with the notion of epistemic privilege focus on arguing that in fact, most of the epistemic goods and epistemic abilities acquired in virtue of experiencing being socially marginalized and oppressed are equally accessible by those who do not undergo such experiences. Briana Toole, a recent example, distinguishes between two notions of epistemic privilege.² First, there is the epistemic privilege afforded by the experiences of those who occupy a given marginalized social location. The kind of epistemic privilege afforded to those who undergo the kind of experiences characterizing the occupation of socially marginalized locations include the possibility of "noticing aspects of the world that are unlikely to be attended to by those who are not marginalized," the development of "certain habits of attention," and "motivations to see more clearly."³ Second, there is the epistemic privilege of an achieved standpoint. Achieving a standpoint, according to Toole, should be understood as a form of *consciousness raising* whereby a group of people with similar experiences get together to conceptualize the wrongs they experience and thus generate knowledge of those wrongs. Toole argues that we should think of the experiences had in virtue of occupying a socially marginalized location as grounding the generation of knowledge found at the standpoint level but without the experiences being sufficient for this knowledge.⁴ In other words, just having the kind of experiences had in virtue of occupying a socially marginalized location does not guarantee that one also acquires knowledge of the wrongs one suffers. For that, one needs to partake in consciousness raising. Moreover, Toole asks us to conceive consciousness raising as a form of training led by experts that can initiate both those who have undergone the kind of experiences had by socially marginalized people and those who have not into knowledge of the wrongs suffered by them. According to Toole, this account

1 Narayan, "Working Together Across Difference," 38; Anderson, "Situated Knowledge and the Interplay of Value Judgments and Evidence in Scientific Inquiry," 502; Wylie, "Feminist Philosophy of Science," 63; and Garcia, *We Are Not Born Submissive*, 87.

2 Toole, "Standpoint Epistemology and Epistemic Peerhood," 2.

3 Toole, "Standpoint Epistemology and Epistemic Peerhood," 3.

4 Toole, "Standpoint Epistemology and Epistemic Peerhood," 8.

has a number of advantages. First, it avoids the implausible claim that simply having the kind of experiences had in virtue of occupying a socially marginalized location automatically guarantees knowledge of the wrongs suffered. Second, it opens up the possibility that those who have not had the relevant kind of experiences of being wronged because of not occupying a socially marginalized location can nevertheless partake in consciousness raising and acquire the same kind of knowledge of the wrongs suffered by those who do undergo the relevant kind of experiences.

Like Toole, Lidal Dror also stresses that the experience of being socially marginalized is often not necessary to acquire an understanding of the moral wrong of social marginalization and oppression. As he puts it in his article "Is There an Epistemic Advantage to Being Oppressed?":

The wrong-making features of various forms of oppression consist largely in objective features about the ways in which people are treated (e.g., unequally or with their freedoms constrained)... Accordingly, claims about oppression and social marginalization are generally objectively analyzable claims about certain structural relations, and are not claims with truth values that are determined by how one person feels about them.⁵

As I argue in section 2 below, Dror's position is more nuanced than this quote makes it seem. Still, a fair characterization of his view is that since the wrong-making features of oppression are objective and do not depend on the experiences had by those who are wronged, epistemic access to them is independent of whether one has the kind of experiences that characterize the occupation of socially marginalized locations. It is instructive that both Toole and Dror focus on the same kind of epistemic abilities. Like Toole, Dror discusses the ability to epistemically justify one's beliefs about oppression (632), "to identify subtler manifestations of oppression" (630), "seeing oppression in new contexts" (631), and the knowledge-how of making good inferences (635). Both Toole and Dror argue that the acquisition of these epistemic abilities in relation to moral wrongs is, at least in principle, independent of whether one experiences being wronged in the relevant way.

Both accounts are partly motivated by the concern to hold accountable those who perpetuate wrongs such as social marginalization and oppression due to their lack of knowledge of how these wrongs work. The thought is that if experience does not play a central role in acquiring knowledge of the relevant

5 Dror, "Is There an Epistemic Advantage to Being Oppressed?" 632. Hereafter, this article is cited parenthetically.

moral wrongs, then those who do not undergo the relevant experiences should be held accountable for their ignorance. As Dror puts it, “precisely because the non-oppressed aren’t in principle significantly epistemically disadvantaged, *they are culpable for the ignorance and lack of care they show towards the existence and functioning of systems of oppression*” (620–21).⁶ Although I agree with both Toole and Dror on this point, this concern can divert attention from a crucial epistemic good and related epistemic ability that can be afforded only to those who experience being wronged in the relevant way, for example, by occupying a socially marginalized location.

It is important to clarify from the start that I do not argue that having the experience of being wronged by, say, being socially marginalized and oppressed puts one in an epistemic privilege vis-à-vis those who do not undergo the same type of experiences. Rather, the aim is to focus attention on an epistemic good and related epistemic ability that have been ignored in the debate. There are two reasons why this is important. First, a corollary of this claim is that a constitutive part of the moral wrongness of, say, oppression is its infliction of the kind of emotional pain mentioned above on those who are wronged. Thus, by articulating the kind of understanding of moral wrongness afforded to those who undergo the relevant experiences of being wronged, we can at the same time shed light on a key aspect of the moral wrongness of the relevant deed. Second, we value people who have undergone experiences of severe moral wrongs and who then commit to teach others about these moral wrongs. We value them, at least in part, as epistemic agents. The argument of this article explains why.

To drive these points home, let me illustrate them by considering Holocaust survivors. Holocaust survivors have played a crucial part in shaping the moral sensitivity of European society since World War II by recounting their experiences of the Holocaust. For example, their testimonies have been part of trials, their books have been taught in schools, and their stories told in interviews and documentaries. What these testimonies bring out is the evil committed during the Holocaust. Now compare the attitudes of those who have not experienced the Holocaust towards, on the one hand, Holocaust survivors and, on the other hand, historians whose expertise is the Holocaust but who were not alive during the Holocaust. Although it is undeniable that the kind of information that historians possess plays an essential role in understanding the kind of moral wrong done in the Holocaust—for example, by making us understand that the Holocaust was, at least in part, the culmination of the widespread antisemitism that characterized European society before the Holocaust—it

6 See also Toole, “Standpoint Epistemology and Epistemic Peerhood,” 4–5.

is not plausible to simply swap the testimony of a Holocaust survivor with the information provided to us by a historian of the Holocaust. How so?

This article argues that the reason is that those who have experienced being wronged in the relevant way are afforded an understanding of the moral wrongness of what they experienced that others cannot give voice to because it appeals to what it means to be someone who has had to lead a life characterized if not defined by the emotional and physical pain caused by the relevant experienced events. Applied to Holocaust survivors, this means that we listen to their testimonies at least in part because they give voice to what it means to endure the emotional and physical pain caused by having experienced being imprisoned in concentration camps and what it means to be someone who survived them and whose life is defined by that survival. The corollary of this is that the evil of the Holocaust at least in part consists in having caused this kind of emotional pain to the people it wronged. This is also why their testimonies provide a unique contribution to our understanding of the Holocaust: they were the ones who were wronged, and they can tell us what that evil means for someone who needs to lead a life partly defined by it. If this is on the right lines, then something similar should be generalized to those who are wronged in virtue of occupying a socially marginalized and oppressed location.

2. EPISTEMOLOGICAL AND PERSONAL TRANSFORMATIONS IN THE EXPERIENCE OF BEING WRONGED

In this section, I articulate the kind of epistemic good that can be possessed only by those who experience being wronged in the relevant way—for example, by occupying a socially oppressed and marginalized location. I do so by appealing to the notion of *transformative experience* as popularized by Laurie Paul.⁷ For Paul, transformative experiences involve both an epistemological transformation and a personal transformation. On an epistemological level, Paul argues that the subject of experience, in having the experience of, say, being the target of a horrific physical attack, acquires an understanding of what it is like to have the experience.⁸ In virtue of this acquisition, experiences have what Paul calls *subjective values*.⁹ According to Paul, the subjective value of the experience refers to the revelation of what it is like to have the relevant experience, a revelation that can be enjoyed only by having the relevant experience. Importantly,

7 Paul, *Transformative Experience*.

8 Paul, *Transformative Experience*, 16. Paul talks about knowledge rather than understanding. I assume that I can reinterpret Paul's claims for my own purposes by substituting 'understanding' for 'knowledge'.

9 Paul, *Transformative Experience*, 12.

Paul is careful in articulating her claim in terms of types of experiences, not merely token experiences. Paul is not merely arguing that each token experience is different, and therefore, in having a new token experience, one acquires an understanding of what it is like to have that token experience. Rather, Paul argues that in having a new type of experience, one acquires an understanding of what it is like to have that type of experience. In having a transformative experience, then, one is revealed with the nature of the type of experience one is having, and in virtue of this, one acquires an understanding of what it is like to have that type of experience.¹⁰

For the purposes of this article, it is important to avoid an “internalist” interpretation of Paul’s concept of transformative experience.¹¹ The reason why this is important becomes apparent in section 3. To anticipate, I argue that by avoiding such interpretation, one can appeal to an inextricable normative relation between the experience of being wronged and objective wrong-making features of the experienced event. In turn, this is crucial when arguing that the kind of understanding that is afforded by the experience of being wronged is an understanding not merely of features of the experience but also of the object of experience. So let me explain what an internalist interpretation is and how to avoid it. In a number of places, Paul seems to suggest that the subjective value of an experience comes down to being revealed with properties of experience conceived independently of properties of the object of experience. That is what philosophers of perception would refer to as *qualia*—that is, intrinsic, nonintentional properties of experience that constitute its phenomenal character. For example, when Paul discusses Frank Jackson’s thought experiment of Mary the color scientist, Paul follows Jackson in arguing that the understanding that Mary acquires once she sees a red rose for the first time is what red looks like in experience.¹² This is easily interpreted, as Jackson does, as Mary acquiring an understanding of a nonintentional, intrinsic property of the experience—that is, the *qualia* associated with redness. A stronger interpretation that is accepted by contemporary philosophers of perception that distance themselves from notions of *qualia* is that Mary, in seeing a red rose for the first time, does not merely acquire an understanding of a property of the experience. Rather, she also acquires an understanding of a property of the object of experience—that is, what it means for an object to look red. Importantly, the idea is that these two

10 Paul, *Transformative Experience*, 10–14.

11 See the exchange between Campbell (review of *Transformative Experience*) and Paul (“Transformative Choice”), where the former imputes an internalist interpretation to the latter. I thank an anonymous reviewer for pointing out that Paul in fact does not make the mistake Campbell imputes to her.

12 Paul, *Transformative Experience*, 8–12.

acquisitions are connected: one acquires an understanding of a property of the object of experience—for instance, what it means for an object to look red—in virtue of acquiring an understanding of what it is like to experience redness.¹³

Now, according to an internalist interpretation of the subjective value of an experience, an epistemological transformation involves acquiring an understanding of what it is like to have an experience, where this is understood in terms acquiring an understanding of a given qualia—that is, an intrinsic, nonintentional property of experience. The problem, as anticipated above, is that the internalist interpretation fails to reach out to properties of the objects of experience. This problem becomes even more pressing when we move from examples of experiencing color to paradigmatic examples given by Paul of transformative experiences. For example, according to an internalist interpretation, in having a child for the first time, one acquires an understanding of what it is like to have a child, where this is restricted to the qualia of the experience. The internalist interpretation is problematic because by interpreting the epistemic role of what it is like to have a new type of experience in terms of the revelation of properties of the experience, conceived independently of properties of the object of experience, it forgoes the right to appeal to properties of the object of experience to make intelligible what it is like to have the new type of experience. For instance, the internalist interpretation foregoes the right to appeal to, say, the preciousness of a newborn to make intelligible what it is like to have a child for the first time. It then becomes mysterious how one ought to go about making intelligible, for example, what it is like to have a child for the first time. Surely, in attempting to make intelligible what it is like to experience having a child for the first time to those who do not have this experience, one would naturally appeal to properties of the objects of experience—for example, the preciousness of the newborn.

Paul is clear that she does not take an internalist line.¹⁴ For this reason, we should interpret the subjective value of an experience in noninternalist terms. By conceiving the epistemological transformation as such, we are in the position of capturing the idea that in virtue of having the relevant transformative experience, one acquires an understanding not merely of properties of the type

13 This stronger interpretation is accepted by both intentionalists and relationalists. The debate between these two sides is how best to capture this interpretation, either in terms of representational properties of the experience or in terms of a nonrepresentational psychological relation between the subject and the object in which mind-independent objects and their sensory properties partly constituting the experience. For a clear exposition of the debate, see Soteriou, *The Mind's Construction*, ch. 2; and Campbell and Cassam, *Berkeley's Puzzle*.

14 Paul, "Transformative Choice."

of experience one has but also of properties of the object that is experienced. In other words, experience provides us with epistemic access to the properties of the object of experience, not just to properties of the experience itself.

Transformative experiences transform one also on a personal level. How should we think of personal transformation? Paul writes about changes in “what it is like for you to be you,” in the kind of person you “take yourself to be,” and in “revising how you experience yourself.”¹⁵ A fruitful way of thinking about these changes is by appealing to what I call *autobiographical reflection*. This is the ability we exercise when reflecting on our experiences and organizing them into personal histories and thus senses of self.¹⁶ Autobiographical reflection is responsible for what one might term one’s self-image or how one thinks of oneself. I go into more detail in what is involved in autobiographical reflection in section 3. For now, I want to suggest that by appealing to autobiographical reflection, we can understand a transformative experience as involving a personal transformation when the exercise of reflecting on the relevant transformative experience leads to a disruption in the otherwise coherent construction of one’s sense of self. This conception of personal transformation gains strength precisely when considering experiences of being morally wronged as one is when occupying a socially marginalized and oppressed location. When the experience one must make sense of as part of one’s autobiography is the experience of being morally wronged in this way, it involves emotional pain. Emotional pain is caused not only while the wrong is being suffered. It is caused also in the act of recollecting the experience of being wronged. Within the act of recollection, emotional pain is partly caused by the act of trying to accept and make sense of the experience of being wronged as part of one’s autobiography. It hurts to have to reconstruct one’s sense of self as someone who has been wronged in the relevant way. Since this interpretation of personal transformation in experiences of being morally wronged becomes crucial in articulating the distinctive kind of understanding of moral wrongness afforded to those who experience being wronged in the relevant way, I want to substantiate it via the following example of sexual harassment.¹⁷

In her book *Epistemic Injustice*, Miranda Fricker quotes from Susan Brownmiller the story of Carmita Wood, an administrator at Cornell University who was sexually harassed, indeed assaulted, by an academic member of staff:

15 Paul, *Transformative Experience*, 16.

16 Autobiographical reflection is exercised in what both philosophers and developmental psychologists refer to as autobiographical memory. See, e.g., Schechtman, *Staying Alive*; and Fivush, *Family Narratives and the Development of an Autobiographical Self*.

17 I thank an anonymous reviewer for pointing out another powerful example—namely, Susan Brison’s memoir of surviving an extremely violent sexual assault (*Aftermath*).

As Wood told the story, the eminent man would jiggle his crotch when he stood near her desk and looked at his mail, or he'd deliberately brush against her breasts while reaching for some papers. One night as the lab workers were leaving their annual Christmas party, he cornered her in the elevator and planted some unwanted kisses on her mouth. After the Christmas party incident, Carmita Wood went out of her way to use the stairs in the lab building in order to avoid a repeat encounter, but the stress of the furtive molestations and her efforts to keep the scientist at a distance while maintaining cordial relations with his wife, whom she liked, brought on a host of physical symptoms. Wood developed chronic back and neck pains. Her right thumb tingled and grew numb. She requested a transfer to another department, and when it didn't come through, she quit. She walked out the door and went to Florida for some rest and recuperation. Upon her return she applied for unemployment insurance. When the claims investigator asked why she had left her job after eight years, Wood was at a loss to describe the hateful episodes. She was ashamed and embarrassed. Under prodding—the blank on the form needed to be filled in—she answered that her reasons had been personal. Her claim for unemployment benefits was denied.¹⁸

This quote describes both the kind of experience that Wood underwent during sexual harassment and the kind of experiences Wood underwent in the aftermath of being sexually harassed. For example, Brownmiller reports that Wood, among other things, developed physical pain, took time off work, wanted to quit her job, and felt stress at the thought of encountering her harasser. These are all experiences involving both emotional and physical pain. Although it is less explicit in the passage, it is not far-fetched to assume that Wood also continued to undergo emotional pain in attempting to make sense of the experience of being sexually harassed as an event that is part of her autobiography. Before being sexually harassed, Wood did not have to make sense of this experience as part of her autobiography or sense of self. After the experience, she did. This in turn involves a disruption of the sense of self that Wood constructed preceding the experience of being sexually harassed. I surmise that Wood underwent emotional pain not only in recollecting what it was like to undergo being sexual harassed at the time of being sexually harassed but also in the act of having to reconstruct her sense of self as someone who has undergone that experience—that is, as someone who has been sexually harassed.

What I want to suggest is that the emotional pain one has to endure in making sense of the given experience of being morally wronged as part of one's

18 Brownmiller, *In Our Time* (quoted in Fricker, *Epistemic Injustice*, 150–51).

autobiography is itself part of the moral wrongness of the act that one experiences. For example, part of the moral wrongness of sexual harassment is that the person who is sexually harassed has to endure the emotional pain of reconstructing their autobiography and thus sense of self as someone who has been sexually harassed. To put it the other way, the aspect of the moral wrongness of sexual harassment that is epistemically accessed by those like Wood who experienced being wronged in the way she did in being sexually harassed is the emotional pain that is endured in her attempt to make sense of the experience as part of her autobiography.¹⁹

Notice how personal and epistemological transformations are inextricably linked in these kinds of transformative experiences. Recall that an epistemological transformation involves acquiring an understanding not only of properties of experience but also of properties of the object of experience. What I am suggesting is that in the experience of being wronged as one is when sexually harassed, one's undergoing a personal transformation—that is, undergoing the emotional pain of having to reconstruct one's autobiography and sense of self as someone who has been sexually harassed—is partly constitutive of the epistemological transformation one undergoes—that is, the epistemic access to a property of the object of experience, in this case, an aspect of the moral wrongness of sexual harassment. Although I develop further this claim in section 3, we are now in a position to articulate the distinctive kind of understanding afforded to those who experience being wronged in the relevant ways. This distinctive kind of understanding is constituted by an understanding of what it is like to endure the emotional pain of making sense of the given experience of being morally wronged as part of one's autobiography and thus sense of self. One key point here is that the relevant emotional pain is not present simply in making sense of a single experience. Rather, it emerges from the attempt to reconstruct one's sense of self by making sense of how that single experience fits in the whole. In the real-life example of Carmita Wood, she attempted to reconstruct her sense of self by making sense of how the experience of sexual harassment fit in her image of herself. She then could appeal to her understanding of what it is like to be someone who must reconstruct her autobiography and thus sense of self as someone who has been sexually harassed. This

19 I qualify my claim by referring to *an aspect* of the moral wrongness of sexual harassment in order to avoid the claim that undergoing a personal transformation is a form of "acquaintance" with "the essence" of the moral wrongness of sexual harassment in an analogous way that some philosophers of perception argue that seeing what red looks like in experience involves becoming acquainted with the essence of redness. See, e.g., Johnston, "How to Speak of the Colors"; and Lord, "How to Learn About Aesthetics and Morality Through Acquaintance and Deference."

is an understanding of an aspect of the moral wrongness of sexual harassment because having to endure that kind of emotional pain is itself part of why sexual harassment is morally wrong. We are now also in a position to explain why those who do not undergo the experience of being wronged in the relevant way cannot access in the same way the distinctive kind of understanding that is afforded to those who do undergo the relevant type of experience. The reason is that the distinctive kind of understanding afforded to those who undergo the relevant type of experience is an understanding of what it is like to endure the emotional pain of someone who must make sense of one's autobiography and thus sense of self as a person who has been wronged in the relevant way.

To be sure, this does not mean that those who do not undergo the relevant type of experience cannot acquire any kind of understanding of what that sort of emotional pain is like. For example, people who do not undergo the relevant type of experience can listen to people who did. But as Yuri Cath argues, we should think of the understanding of what it is like to have an experience possessed by those who do not undergo the relevant experience in terms of "gradability" that stops short of coinciding with the understanding possessed by those who do undergo the relevant experience.²⁰ How far one thinks one can acquire this kind of understanding depends, at least in part, on how far one thinks that imaginative understanding and empathy can go.²¹ In section 5, I describe in more detail the relationship between the kind of understanding possessed by those who experience being wronged in the relevant ways and the epistemic abilities possessed by those who do not undergo the relevant type of experiences.

3. THE ROLE OF AUTOBIOGRAPHICAL REFLECTION IN GENERATING MORAL CONCEPTS

In this section, I spell out the distinctive epistemic ability that is both acquired and exercised in generating the distinctive kind of understanding of moral wrongness articulated in section 2. I do so by focusing on the role of autobiographical reflection in generating moral concepts. I begin by distinguishing between generating, acquiring, possessing, and deploying a moral concept.

20 Cath, "Knowing What It Is Like and Testimony."

21 For a recent and strong defense of empathic, imaginative perspective-taking abilities in the acquisition of moral understanding, see Bailey, "Empathy and the Value of Humane Understanding." For a relevant discussion of imaginative understanding versus scientific understanding, see Campbell, *Causation in Psychology*. For a position closer to mine that argues that experience influences and limits in important ways our imaginative perspective taking in moral matters, see Toole, "Demarginalizing Standpoint Epistemology."

Once again, I appeal to Fricker's discussion of sexual harassment, more specifically *hermeneutical injustice*.²² Hermeneutical injustice is the wrong done to people who experience, say, sexual harassment but who do not have the epistemic means to label and conceptualize the wrong they experience. Fricker argues this was the case for many women before the advent of second-wave feminism. Because of this, Fricker argues, women who experienced sexual harassment were at a "cognitive disadvantage" that prevented them from understanding the experiences they underwent. For example, according to Fricker, Wood did not have the conceptual resources to articulate her experience. This meant that when she reflected on her experiences, although she endured the relevant type of emotional pain, she was not in a position to say that what she experienced was sexual harassment, and she might not have been in a position even to say that she had been morally wronged. One way in which we can read Fricker is as saying that before second-wave feminism, the concept of sexual harassment had not been generated—that is, although women experienced events that fall under the description of sexual harassment, there was no labeled concept of sexual harassment that they could appeal to in order to describe their experiences. Since the concept of sexual harassment had not been generated, women who experienced sexual harassment could not acquire the concept of sexual harassment to describe their experience. In turn, they could not possess the concept of sexual harassment, and therefore they could not deploy it to describe their experiences. Thus, the generation of a moral concept is foundational for its acquisition, possession, and deployment.

How is a moral concept like sexual harassment generated in the first place? According to a genealogical account shared by David Wiggins, Philip Pettit, and Fricker, the generation of moral concepts begins with the fact that some people undergo certain experiences in response to certain events.²³ These experiences are typically affective, and they involve an event that becomes of concern to the subject of experience. Individually, these people do not have the conceptual resources to make sense of the experience or to generate concepts that can help them do so. It is only once people who have similar experiences get together to constitute a social process with the aim of making sense of their experiences that the generation of the relevant moral concepts begins. The central aim of this social process is the shared effort to make sense of the normative relation between the relevant experiences and what the experiences respond to.

22 Fricker, *Epistemic Injustice*, 150–52.

23 See Wiggins, *A Sensible Subjectivism?*; Pettit, "Realism and Response-Dependence"; and Fricker, *Epistemic Injustice*. See also Fricker, "Epistemic Oppression and Epistemic Privilege"; Pohlhaus, "Relational Knowing and Epistemic Injustice"; and Toole, "From Standpoint Epistemology to Epistemic Oppression."

Eventually, by focusing on this normative relation, the participants start noticing similar features to each other's experiences and to what the experiences are responding. In turn, the process of identifying these similarities leads to a generalization of the type of experience had by the participants and what the type of experience is a response to. The generalization leads to the generation of the relevant moral concept—for example, sexual harassment. The concept of sexual harassment was generated to refer to objective features of experienced events that have a normative relation to the relevant experiences. Once the concept of sexual harassment was generated, the concept became available for acquisition both by those who had experienced sexual harassment and by those who did not. Women who experience sexual harassment can now use it to describe and understand their experiences. Those who have not experienced sexual harassment can also acquire the concept of sexual harassment and use it to explain its moral wrong or to identify manifestations of it.

There are two aspects of what it is to elaborate the normative relation between one's experience of, say, being sexually harassed and the objective features of the experienced event that need spelling out. First, as a number of philosophers of emotion have argued, emotional experiences involve evaluations of events they respond to.²⁴ Using a simplified example, feeling angry at a remark involves an evaluation of the remark as, say, offensive. The evaluation at play in an emotional experience can be either justified or not. For example, if the remark turns out not to have been offensive after all, then one typically changes one's emotional response. In virtue of involving evaluations, then, emotional experiences are open to rational assessment—that is, we give reasons to justify our emotional responses, and doing so involves picking on objective features of the event we emotionally respond to, for example, the words used in the remark. But—and this is crucial—the objective features of the experienced event provide reasons in the justification of an emotional response only because they fall under an evaluative description. The words used in a remark provide one with a reason to justify one's anger only if the words fall under an evaluative description, such as “offensive.” Whether the words do indeed fall under the relevant evaluative description is open to debate. This is to be expected since emotional experiences are open to rational assessment. Applying this to the case at hand, the point is that unpacking the normative relation between an emotional experience and the event it is a response to involves deliberating

24 For accounts exploring the normative relation between experiences of an emotional kind and descriptive features of the objects of experience, see D'Arms and Jacobson, “The Moralistic Fallacy”; Deonna and Teroni, *The Emotions* and “Emotions and Their Correctness Conditions”; and Tappolet, *Emotions, Value, and Agency*. These accounts are inspired by Wiggins, “Sensible Subjectivism?”

on how one's experience fits with objective features of the experienced event, where what is debated is whether the objective features fall under the relevant evaluative description. The social process leading to the generation of a moral concept can then be seen as a social deliberative attempt to generate new conceptual resources to spell out the evaluative description under which objective features of an experienced event can be seen as providing reasons for the justification of the relevant emotional experience. It is of vital importance to appreciate that spelling out this normative relation involves thinking of the relevant experience and what the experience responds to—namely, the objective features of the experienced events—as inextricably connected.

Take the example of Carmita Wood. Suppose she was trying to make sense of her experiences in a collaborative effort with other women who had similar experiences. They might have described their experiences and then attempted to explain why they felt as they did, where this would have involved appealing to objective features of the situation they experienced. Their making sense of their experiences would have been inextricably connected to appealing to objective features of the experienced event. In Wood's example, she described the strength used by the harasser to corner her when forcing kisses on her. Wood also described the new behaviors she adopted as a reaction to the sexual harassment, such as leaving her job. Descriptions of this sort are descriptions of objective features of situations. At the same time, they are descriptions emerging from evaluations of the relevant situations. For instance, Wood's descriptions of how she left her job in the aftermath of what happened to her emerged from the evaluation of the wrongness of what happened to her. Leaving her job is an aspect of the wrongness that she suffered. Descriptions of objective features of situations of the sort given by Wood are the result of evaluations. In turn, these evaluations are part of the experiences had by Wood. The upshot is that neither side can be made intelligible without the other. Leaving her job is made intelligible as a wrong done to her in light of the experiences she underwent. And the descriptions of her experiences are made intelligible by appealing to the objective features of the situations that her experiences were responses to.

The second aspect of what it means to elaborate the normative relation between the relevant experience of being wronged and the experienced event is specific to those who undergo the relevant experience. This can be appreciated by spelling out why autobiographical reflection is an important component of the process of generating moral concepts like sexual harassment. Women who were sexually harassed in the past had to endure the emotional pain of reflecting back on their experiences in the social process described above. That distinguished them from those who also elaborated the normative relation between

the experience of being sexually harassed and the experienced event, thus being part of the social process, but who had not been sexually harassed. Those like Wood who had been sexually harassed exercised an evaluation of the experienced event that differed in at least one important respect from the evaluation exercised by those who did not experience sexual harassment. That is, for those like Wood who were sexually harassed, the attempt to spell out what was being evaluated in the experience of being sexually harassed—i.e., the relevant moral wrong—was inextricably connected to the fact that they were attempting to make sense of the experience as part of their personal history.²⁵ In other words, their evaluation of the wrong suffered in the experienced event was informed by what it means to be someone who needs to lead a life as someone who has been sexually harassed. This involves emotional pain. So, unlike those who had not been sexually harassed, Wood's attempt at spelling out the experienced wrong was not of a detached sort, the kind that a bystander might attempt while listening to Wood's experiences. Moreover, for those like Wood, there is a close relationship between the two aspects just elaborated of what it means to spell out the normative relation between an emotional experience and what it is responding to. For people like Wood, evaluating objective features of experienced events is informed by their attempt at making sense of the experiences as part of their autobiography. That is because they try to conceptualize evaluatively the objective features of the experienced events in a way that justifies the way they feel, which, in turn, involves attempting to make sense of themselves as someone who needs to lead a life characterized by the relevant experience.

What emerges from the above, then, is a distinctive epistemic ability acquired and exercised by those like Wood. That is, the ability to extrapolate a generalizable understanding of the moral wrongness of an experienced event from the personal experiences one has undergone. Within the context of generating moral concepts, this epistemic ability is distinctive to those who undergo the relevant experiences because the understanding of the relevant moral wrong emerges from the effort of piecing together emotionally painful experiences into a sense of self. Unless one has gone through the relevant experiences, one is not in a position to attempt to piece them together into a coherent sense of self. One has to undergo the relevant experiences to be in the position of attempting to make sense of them within one's autobiography. Moreover, and importantly, spelling out this distinctive epistemic ability shows why not everyone who has had the relevant experience automatically acquires the epistemic good articulated in section 2. This is because not everyone is able

25 For an insightful account of the relationship between what it means to evaluate and what it means to lead a life, see Calhoun, *Doing Valuable Time*.

to exercise the epistemic ability articulated in this section. As Avishai Margalit says in relation to Holocaust survivors, “although all sufferers of evil are equal in being qualified to attest to their suffering, they are far from equal in their ability to elucidate their experience of evil to us who were not there. This is a great achievement that should not be scorned because it may offend an alleged democratic instinct about witnesses.”²⁶ The achievement consists in the ability to face one’s emotional pain and spell out the objective features of the experienced event in light of the evaluation exercised in the very act of recollection of the relevant experience.

4. BEING WRONGED AND EPISTEMIC ABILITIES

In this section, I defend the claim made in sections 2 and 3 by replying to the objection that articulating what it is like to be, say, oppressed or sexually harassed does not entail making claims that go beyond the experience itself and onto objective features of these wrongs.²⁷ I do so by replying to a number of claims made in Dror’s argument. To be sure, Dror’s overall target is the “strong inversion thesis,” according to which “socially marginalized people, by virtue of their social location *qua* social location, have a superior epistemic position than non-oppressed people when it comes to knowing things about the workings of social marginalization that concern them” (628). As stated in section 1, my aim is not to defend the strong inversion thesis. Rather, my aim in this section is to defend the view that the epistemic good and ability articulated in sections 2 and 3, respectively, allow one to make claims that go beyond the experience itself to objective features of the given moral wrong.

Dror’s argument is divided into three parts. I consider each in turn. The first part devises an alternative version of Jackson’s thought experiment about Mary the color scientist. In Dror’s version, Jane is a white woman who is an expert in the Black civil rights movement. In virtue of her expertise, Jane possesses all descriptive and normative facts about the oppression of Black people in the United States. At the same time, Jane has never experienced being a Black woman subjected to racial discrimination. Dror concedes that Jane cannot possess an understanding of what it feels like to have the kind of emotions that a Black woman experiences as a response to being oppressed. That is out of Jane’s reach. At the same time, Dror argues that it is implausible to think that therefore Jane cannot possess the same kind of “understanding [of] the operations of racism within the United States” (629). In Dror’s words, “the qualia of being

²⁶ Margalit, *The Ethics of Memory*, 181–82.

²⁷ I thank an anonymous reviewer for putting it this way.

oppressed does not generally give better epistemic support to claims about the workings of social marginalization" (629).

I do not deny that there are many ways in which one can acquire knowledge of the workings of moral wrongs such as oppression that do not rely on experiencing these moral wrongs. Nevertheless, the problem with the first part of Dror's argument is that Dror conceives the phenomenology of experience in terms of qualia. As we have seen in section 2, qualia is a theoretical construct appealed to by some views in the philosophy of perception to capture the phenomenal character of experience. According to these views, the phenomenal character of experience is constituted by nonintentional properties of experience, i.e., qualia. Since qualia are nonintentional properties, appealing to them brings with it the assumption that there is a contingent relation between the phenomenology of experience and the properties of the mind-independent object of experience. In other words, appealing to qualia brings with it the assumption that specifying the phenomenology of experience is independent from specifying the properties of the mind-independent objects the experience is an experience of. This leads one to the "internalist" conception of the phenomenology of experience discussed in section 2. It should not be surprising, then, that Dror appeals to qualia when driving a wedge between knowledge of what it is like to have the experience and knowledge of objective features of events that one experiences.

It might be argued that I assume too much of what is involved in Dror's appeal to qualia. But we find a similar problem in the second part of Dror's argument, where he focuses specifically on emotions. Dror takes issue with Narayan's claim, quoted above, that it is in virtue of the emotional experiences had by oppressed people in response to oppression that they acquire knowledge of oppression that is out of reach to those who are not oppressed. Dror argues that having the relevant emotional experiences does not provide one with the epistemic ability to make claims about the objective features of, say, oppression that are in principle out of reach to those who did not have the relevant experiences. What kind of epistemic abilities does Dror have in mind? As we saw in section 2, Dror lists the ability to epistemically justify one's beliefs about oppression (629), "the ability to identify subtler manifestations of oppression" (630), the ability to "[see] oppression in new contexts" (631), and the knowledge-how of making good inferences (635).

I concede that the above listed epistemic abilities can be accessed by those who do not have the relevant emotional experiences. The problem is that the list is not exhaustive. As argued in section 3, there is an epistemic ability that is available only to those who have undergone the relevant experiences. That is, the ability to extrapolate a generalizable understanding of the moral

wrongness of an experienced event from the personal experiences one underwent. To reiterate, this epistemic ability is available only to those who have undergone the relevant experiences because it involves an evaluation of the experienced events that is inextricably connected to the attempt to restructure one's sense of self as someone who has undergone the relevant experience. In other words, the understanding of the relevant moral wrong emerges from the effort of making sense of emotionally painful experiences as part of one's sense of self. This is itself part of the evaluation of the objective features of the experienced event. Crucially, then, this is not reducible to mere knowledge of what it is like to be oppressed, understood independently of making claims about objective features of oppression. This is because, as we have seen, elaborating the normative status of one's experiences is inextricably connected with appealing to the objective features of the experienced event. One appeals to certain specific objective features of the experienced event in virtue of evaluating them—that is, insofar as one sees those objective features under a given evaluative description. But in turn, the evaluation is part of the experience, of “what it's like” to be oppressed. Of course, this is not to say that one cannot make claims about objective features of oppression without having the relevant experiences. Rather, it is to say that the epistemic ability articulated in section 3 is not confined to making claims about experience conceived independently of the objective features of the experienced event.

Third, one might argue that in fact, Dror concedes that there is an epistemic ability available only to those who undergo experiences of being oppressed (633). Moreover, he even seems to suggest that this epistemic ability is connected to making claims about the normative status of social marginalization. If so, where is the disagreement? This objection arises from Dror's distinction between two types of claims that the socially oppressed can make about social marginalization: those that are made true by how one feels about them and those that are not. According to Dror, “the vast majority of normative and descriptive claims about the workings of social marginalization” are of the latter type—that is, they are not made true by how one feels about them (632). By contrast, Dror argues, claims about social marginalization made by oppressed people that are made true by how they feel about them fall under the *negative affect advantage* of oppressed people (633). The negative affect advantage consists in having epistemic access to “knowing *whether* a particular experience is hurtful” and “to normative and descriptive claims about the workings of social marginalization” (633).

The problem is that Dror is unclear how we are to think of the relationship between “knowing *whether* a particular experience is hurtful” and “normative and descriptive claims about the workings of social marginalization.” For

example, one might think of that relationship in causal terms. For instance, assuming one's emotional experiences are appropriately attuned, then whenever one feels hurt, one knows that the experienced event is morally wrong. This option is compatible with Dror's subsequent claim that "what really matters for the negative affect advantage is that (and perhaps how much) someone was hurt, rather than what exactly the hurt feels like" (633). It is not hard to see why only those who undergo these kinds of experiences have this kind of advantage. All it consists in is the ability to say that an experience is hurtful. I believe this option to be unappealing. It reduces epistemic agents to mere reliable detectors of moral wrongs, like a smoke alarm.²⁸ Moreover, interpreting the relation in causal terms precludes appealing to a normative relation between the experience and the experienced event. If instead Dror does opt for a normative relation, i.e., a relation that allows for justification, then it becomes unclear whether Dror can hold onto his earlier claims made in relation to the analogy between Jane and Mary, where he severs our understanding of features of the experience from our understanding of objective features of the experienced event. This is because, as argued in section 3, postulating a normative relation between the experience of being wronged and the objective wrong-making features of the experienced event commits one to the view that there is an inextricable relation between the two. The upshot is that we find the answer to Dror's objection in the normative relation between the experience of being wronged and the objective wrong-making features of the experienced event that Dror appeals to.

5. THE EDUCATIONAL ROLE OF THE EXPERIENCE OF BEING WRONGED

The discussion in section 4 above raises the question of the relationship between the distinctive kind of epistemic ability I argue is afforded to those who experience being morally wronged in the relevant way and the epistemic abilities that can be acquired and exercised without undergoing experiences of this kind. In this section, I do two things. First, I show why the distinctive kind of epistemic ability afforded to those who experience being wronged in the relevant way is not reducible to and thus must be distinguished from the epistemic abilities that can be acquired by those who do not undergo the relevant experience. Second, I articulate the educational role of the experience of being wronged in relation to the epistemic abilities that can be acquired by those who do not undergo the relevant experience.

28 Johnston, M. "The Authority of Affect."

Recall that Dror mentions the following epistemic abilities that can be acquired by those who do not undergo the experience of being morally wronged in the relevant way: the ability to epistemically justify one's beliefs about oppression, "the ability to identify subtler manifestations of oppression," the ability to "[see] oppression in new contexts," and the knowledge-how of making good inferences. These are reminiscent of the epistemic abilities debated in current moral epistemology. The first and last epistemic abilities mentioned by Dror are reminiscent of the epistemic ability at play in the account of moral understanding made notorious by Alison Hills.²⁹ In a number of papers, Hills argues that we should conceive the epistemic ability that is partly constitutive of moral understanding strictly in terms of a deliberative ability that she calls "understanding why."³⁰ Hills conceives "understanding why" as a deliberative ability in that it involves a facility with connecting propositions in such a way as to enable one to demonstrate why a given proposition is true. If one is able to do so, according to Hills, then one "grasps" the relation between the relevant proposition and the reasons that make it true: "Moral understanding involves a grasp of the relation between a moral proposition and the reasons why it is true."³¹ In turn, in virtue of such "grasping," the epistemic agent possesses *cognitive control* over them. Hills fleshes out the notion cognitive control through the notion of "manipulation": "if you understand why *p* (and *q* is why *p*) then you have cognitive control over *p* and *q*, and thus you can (in the right circumstances) manipulate the relationship between *p* and *q*."³² Hills closely links the deliberative ability that characterizes "understanding why" with explanation. If one understands why a given moral proposition is true, then one is able to explain why a given proposition is related in the appropriate way to other propositions and how together they ground the truth of the relevant proposition.

The second and third epistemic abilities mentioned by Dror are reminiscent of epistemic abilities appealed to by philosophers who argue that our moral sensibility should be conceived in terms of perception-like experiences.

29 See Hills, "Moral Testimony and Moral Epistemology" and "Understanding Why." See also Simion, "The Explanation Proffering Norm of Moral Assertion"; Lewis, "The Norm of Moral Assertion"; Kelp, "Moral Assertion"; Croce, "Moral Understanding, Testimony, and Moral Exemplarity"; Boyd, "Moral Understanding and Cooperative Testimony"; Hills, "Moral Testimony"; and Malfatti, "Can Testimony Transmit Understanding?" and "On Understanding and Testimony."

30 Hills, "Moral Testimony and Moral Epistemology" and "Understanding Why."

31 Hills, "Moral Testimony and Moral Epistemology," 101.

32 Hills, "Understanding Why," 663.

Paulina Sliwa is a recent example.³³ Sliwa explicitly reacts against Hills, complaining that Hills's account of moral understanding entails that we can arrive at moral conclusions only if deliberation is involved. By contrast, for Sliwa, perception-like experiences in which one exercises one's knowing right from wrong are instances in which we can exercise our moral understanding without deliberation. Since Hills conceives moral understanding as constituted by deliberation, Sliwa argues, it cannot capture the special relationship between, on the one hand, perception-like experience and, on the other hand, the acquisition and exercise of moral understanding. Sliwa illustrates her argument by quoting from George Orwell's "A Hanging," where Orwell describes witnessing a public execution.³⁴ Sliwa argues, "It is natural to say that witnessing the execution led Orwell to understand that capital punishment is morally wrong and why it's wrong. . . . Orwell's moral insight is not based on moral deliberation. He didn't reason his way to the conclusion that the death penalty is morally wrong. Rather, it is based on something more like a perceptual experience. He saw 'the unspeakable wrongness' of killing another human being."³⁵

The understanding of moral wrongness that I argue is afforded by the experience of being wronged is not reducible either to the deliberative ability to explain the truth of a moral proposition or the perception-like ability to identify morally relevant features of the object of experience. In the case of deliberation, this is most apparent at the level of acquisition. Hills is clear that experiencing a given moral wrong, either as a moral wrong being done to someone or of being wronged oneself, is not necessary to acquire the deliberative ability to explain the truth of a moral proposition.³⁶ All that is needed to acquire the ability to explain the truth of a moral proposition is the acquisition of the relevant propo-

33 See Sliwa, "Moral Understanding as Knowing Right from Wrong." The notion of perceptual-like experience of moral features of our environment has been popularized by John McDowell's work on virtue ethics. See in particular McDowell, "Virtue and Reason." McNaughton (*Moral Vision*) and Dancy (*Moral Reasons*) defend a version of McDowell's account. For a recent collection of papers on the topic, see Bergqvist and Cowan, *Evaluative Perception*. A number of philosophers suggest that the kind of perception-like experiences that detect moral properties of objects and events should be conceived as affective in kind. The account below applies to them too. See, for example, Johnston, "The Authority of Affect"; Zagzebski, "Emotion and Moral Judgement"; Doering, "Seeing What to Do"; and more recently, Cowan, "Epistemic Perceptualism and Neo-Sentimentalist Objections"; Montague, *The Given*; Tappolet, *Emotions, Value, and Agency*; Lord, "How to Learn About Aesthetics and Morality Through Acquaintance and Deference"; and Poellner, *Value in Modernity*.

34 Orwell, "A Hanging," 69.

35 Sliwa, "Moral Understanding as Knowing Right from Wrong," 544–45.

36 Hills, "Moral Testimony," 411.

sitions and the ability to reason with the relevant propositions. This applies also at the level of exercising the given understanding. At the level of exercise, Hills argues that moral understanding is exercised strictly in the ability to reason to the truth of a given moral proposition. Experience does not feature in any way at the level of exercise of moral understanding. Sticking to the level of acquisition, the kind of understanding of moral wrongness articulated in sections 2 and 3 necessitates the experience of being wronged to be acquired. To reiterate, the claim is that the experience of being wronged affords an understanding of what it is like to endure the emotional pain of having to make sense of the given experience as part of one's autobiography and thus sense of self. The assumption here is that part of the moral wrongness of, say, sexual harassment is that it inflicts on the person who is sexually harassed, among other things, the emotional pain of having to make sense of the experience of having been sexually harassed as part of their autobiography and thus sense of self. This kind of understanding is not reducible to the kind focused on by Hills because of its constitutive relationship to the experience of being wronged.

The understanding of moral wrongness that I argue is afforded by the experience of being wronged is also not reducible to the perception-like ability to identify the moral wrongness of a given situation. The reason is that it is not necessary to have the kind of perception-like experience focused on by Sliwa that the moral wrong is suffered by the one having the experience. That is, one need not be morally wronged to perceptually identify the moral wrongness of a situation. For example, suppose that someone who has never been sexually harassed is on the bus when they suddenly see a man rubbing himself against a woman. Suppose also that they immediately call out the act and take appropriate action. In this case, Sliwa would be right to say that they did not reason their way to the conclusion that what the man did was morally wrong (to put it mildly). Rather, they saw, or quasi-perceptually identified, the moral wrongness of what the man did—more specifically, that the man was sexually harassing the woman. What this example makes clear is that the perception-like experience of moral wrongness had by those who have never experienced being wronged in the relevant way can be had by bystanders who witness the moral wrong being done to someone else. Having this perception-like experience does not necessitate that the moral wrong is done to them. This is the defining difference with the kind of experience had by someone when they are sexually harassed. They are not a mere bystander. The moral wrong is done to them, and this is reflected in the kind of understanding that I argue is afforded by the experience of being wronged.

I now turn to the educational role of the experience of being wronged. The key point of contact between the distinctive kind of understanding that I argue

is afforded to those who experience being wronged and the notion of moral understanding defended by Hills is the following. One key ability upon which deliberation of this sort relies is one's understanding of the relevant moral concept—that is, what the relevant moral concept refers to. For example, one's ability to explain why sexual harassment is morally wrong depends on one possessing an understanding of what the concept of sexual harassment refers to. This is perhaps most evident when we consider, as Hills rightly argues, that the ability to explain the truth of moral propositions involves the ability to appeal to reasons and to manipulate reasons in a way that it builds the explanation. In other words, the ability to explain why sexual harassment is morally wrong involves appealing to reasons why this is so. Now, there are a variety of reasons that people who have not experienced sexual harassment appeal to in order to explain why it is morally wrong. Among these are reasons that emerge from the descriptions of the experience of being sexual harassed. This happens in two ways. First and more directly, in their description of what it means to be sexually harassed, one might articulate reasons why sexual harassment is morally wrong. People who have never experienced being sexually harassed can then appeal to these reasons in explaining why sexual harassment is morally wrong. Second and indirectly, those who never had the relevant kind of experience can exercise imaginative perspective taking when listening to those who have when they recount what it is like to be sexually harassed and, in so doing, come up with new reasons why sexual harassment is morally wrong.³⁷ To be sure, there are important limits to how far one's imaginative perspective taking can go and thus how far one's ability to come up with new reasons justifying beliefs can stretch. Nevertheless, listening to people recount their experiences of being morally wronged in the relevant way can facilitate one's ability to explain the truth of moral propositions, such as that sexual harassment is morally wrong.³⁸

37 I owe this observation to Toole, "Demarginalizing Standpoint Epistemology." Toole appeals to Paul's account of "cognitive empathy" to argue that imaginative perspective taking can lead to the acquisition of *de se* knowledge, defined as "personal knowledge that one expresses or grasps using first-personal concepts, e.g. 'I,' 'me,' 'mine,' and so on" (13). See also Paul, "First Personal Modes of Presentation and the Structure of Empathy." Toole argues that *de se* knowledge can grant epistemic access to evidence justifying beliefs that would otherwise not be accessible.

38 The above applies also to accounts of moral understanding that demand a stronger place for our affective and motivational dispositions because these accounts agree with Hills that the cognitive component of moral understanding is the ability to appeal to reasons in explaining the truth of a moral proposition. See Enoch, "A Defense of Moral Deference"; Howell, "Google Morals, Virtue, and the Asymmetry of Difference"; Fletcher, "Moral Testimony"; and Callahan, "Moral Testimony."

Just like in the case of moral deliberation, people who experience being morally wronged in the relevant way can help those who have not had the relevant experiences develop the perception-like ability to identify features of the relevant moral wrong. This becomes apparent when we consider accounts that draw an analogy between ethical virtue and practical skills.³⁹ Assuming that ethical virtue involves a perception-like sensibility to features of situations that are relevant to the moral rightness or wrongness of what one experiences, the idea is that one can acquire this kind of perception-like sensibility in an analogous way as one can learn a practical skill.⁴⁰

I want to suggest that people who not only have been wronged in the relevant ways but also have reflected on their experiences in ways that can contribute to shaping the relevant moral concepts should be considered as eligible moral trainers for those who want to learn how to perceptually identify morally relevant features of situations. The reason is that their testimonies are in effect giving voice to their evaluation of objective features of the events they experienced. In turn, those who have not experienced the relevant wrongs can avail themselves of these evaluations to detect manifestations of the same wrong in new situations. Within this context, it helps to think of perception-like sensibility as a faculty that, as Mark Johnston puts it, can be “refined.”⁴¹ As Johnston argues, perception-like sensibility is refined through acts of imagination and affective responses. Here, the testimonies of those who have undergone the relevant experiences provide the platform for these refinements for those who have not undergone the relevant experiences.

Recall the example of Holocaust survivors from section 2. It is often the case that Holocaust survivors recount their experiences to identify objective features of contemporary societies that are reminiscent of oncoming authoritarianism. The most recent example is a group of Holocaust survivors who urged voters before the June 2024 European elections to vote against what they recognized as rising right-wing political parties.⁴² The same eligibility to be moral trainers should be generalized to those who are able to articulate

39 See Annas, *Intelligent Virtue*; Dougherty, “The Importance of Roles in the Skill Analogy”; Fridland, “Motor Skill and Moral Virtue”; Fridland and Stichter, “It Just Feels Right”; Jacobson, “Seeing by Feeling”; Stichter, “Virtues as Skills, and the Virtues of Self-Regulation”; and Swartwood, “Wisdom as an Expert Skill.”

40 See McDowell, “Virtue and Reason”; Jacobsen, “Seeing by Feeling”; and Fridland, “Motor Skill and Moral Virtue.”

41 Johnston, “The Authority of Affect,” 206.

42 Connolly, “Holocaust Survivors Urge Young Europeans to Vote Against Far Right.” *Guardian*, June 5, 2024, <https://www.theguardian.com/world/article/2024/jun/05/holocaust-survivors-young-europeans-vote-against-far-right-eu-election>.

their evaluations of objective features of, for example, oppression and sexual harassment in virtue of understanding what it is like to lead a life characterized by having experienced these wrongs. By reflecting on their experience of being sexually harassed, people who have been wronged in the relevant way can identify features of the actions of the perpetrator they endured that justify how they felt at the time of the experience and how they feel at the time of the recollection of the experience. That is, they articulate their evaluation of the experienced event. In turn, this puts them in the position to teach others how to identify the relevant descriptive features of a situation that fall under the concept of sexual harassment.

Two points of clarification. First, I am not denying that people who have not undergone the relevant experiences can be eligible moral trainers. Rather, I am arguing that those who have experienced the relevant moral wrongs can offer unique lessons to those who have not experienced the relevant wrongs. That is because they give voice to a specific kind of evaluation of the experienced events, and this can be the basis for developing the epistemic abilities focused on by Toole and Dror. Second, it is important to distinguish moral trainers as described here from some conceptions of moral experts. Specifically, Julia Driver argues that moral experts make consistently better moral judgements, and they possess a better understanding of morality as a whole.⁴³ This does not necessarily apply to those whose moral teaching is based on their experience of being morally wronged. Those who generate the distinctive kind of understanding articulated in section 2 have a unique understanding of the relevant moral wrong in virtue of understanding what it is like to lead a life characterized by the relevant experiences. This means that their understanding is, at least at first, limited to specific moral wrongs—for example, sexual harassment. Nothing I say in this article suggests that this kind of understanding overreaches to other moral wrongs, let alone morality as a whole. More is needed in terms of argumentation to show that this kind of understanding overreaches in these ways.

6. CONCLUSION

I conclude by spelling out one upshot of my argument. That is, what the above shows is that the account of this article does not corrode the moral accountability of those who have not experienced being morally wronged in the relevant way and who are therefore not afforded the same kind of understanding. Not having experienced the relevant moral wrong and thus not being afforded the distinctive kind of understanding that is afforded to those who have

43 Driver, "Moral Expertise."

experienced being morally wronged do not exculpate one for relevant moral failures. On the contrary, the account of this article emphasizes that there are a number of valuable epistemic abilities that those who have not experienced being morally wronged in the relevant way can nevertheless acquire and that therefore they are morally accountable for their actions. Not only that, but the account here goes a step further by arguing that an important way in which people who have not experienced being morally wronged in the relevant way can acquire the epistemic abilities that make them morally accountable is by being taught by those who have been morally wronged in the relevant way. In these cases, people who have not experienced being morally wronged in the relevant way owe their ability to understand the moral wrongness of, say, sexual harassment to those who have. For this reason, people who have been morally wronged in the relevant way and who have exerted conceptual effort in generating an understanding of the moral wrongness of what they experienced should be seen as potential moral educators.

University College London
d.vanello@ucl.ac.uk

REFERENCES

- Anderson, Elizabeth. "Situated Knowledge and the Interplay of Value Judgments and Evidence in Scientific Inquiry." In *In the Scope of Logic, Methodology and Philosophy of Science*, edited by P. Gärdenfors, J. Woleski, and K. Kijania-Placek. Kluwer, 2012.
- Annas, Julia. *Intelligent Virtue*. Oxford University Press, 2011.
- Bailey, Olivia. "Empathy and the Value of Humane Understanding." *Philosophy and Phenomenological Research* 104, no. 1 (2020): 50–65.
- Bergqvist, Anna, and Robert Cowan, eds. *Evaluative Perception*. Oxford University Press, 2018.
- Boyd, Kenneth. "Moral Understanding and Cooperative Testimony." *Canadian Journal of Philosophy* 50, no. 1 (2020): 18–33.
- Brison, Susan. *Aftermath: Violence and the Remaking of a Self*. Princeton University Press, 2023.
- Brownmiller, Susan. *In Our Time: Memoir of a Revolution*. Dial Press, 1990.
- Calhoun, Cheshire. *Doing Valuable Time*. Oxford University Press, 2018.
- Callahan, L. F. "Moral Testimony: A Re-Conceived Understanding Explanation." *Philosophical Quarterly* 68, no. 272 (2018): 437–59.
- Campbell, John. *Causation in Psychology*. Harvard University Press, 2020.

- . Review of *Transformative Experience*, by L. A. Paul. *Philosophy and Phenomenological Research* 91, no. 3 (2015): 787–93.
- Campbell, John, and Quassim Cassam. *Berkeley's Puzzle*. Oxford University Press, 2014.
- Cath, Yuri. "Knowing What It Is Like and Testimony." *Australasian Journal of Philosophy* 97, no. 1 (2019): 105–20.
- Cowan, Robert. "Epistemic Perceptualism and Neo-Sentimentalist Objections." *Canadian Journal of Philosophy* 46, no. 1 (2016): 59–81.
- Croce, Michel. "Moral Understanding, Testimony, and Moral Exemplarity." *Ethical Theory and Moral Practice* 23, no. 2 (2020): 373–89.
- Dancy, Jonathan. *Moral Reasons*. Blackwell, 1993.
- D'Arms, Justin, and Daniel Jacobson. "The Moralistic Fallacy: On the 'Appropriateness' of Emotions." *Philosophy and Phenomenological Research* 61, no. 1 (2000): 65–90.
- Deonna, Julien, and Fabrice Teroni. *The Emotions: A Philosophical Introduction*. Routledge, 2012.
- . "Emotions and Their Correctness Conditions: A Defense of Attitudinalism." *Erkenntnis* 89 (2024): 45–64.
- Döring, Sabine A. "Seeing What to Do: Affective Perception and Rational Motivation." *Dialectica* 61, no. 3 (2007): 363–94.
- Dougherty, Matthew Ryan. "The Importance of Roles in the Skill Analogy." *Journal of Ethics and Social Philosophy* 17, no. 1 (2020): 75–102.
- Driver, Julia. "Moral Expertise: Judgment, Practice, and Analysis." *Social Philosophy and Policy* 30, nos. 1–2 (2013): 280–96.
- Dror, Lidal. "Is There an Epistemic Advantage to Being Oppressed?" *Nous* 57, no. 3 (2023): 618–40.
- Enoch, David. "A Defense of Moral Deference." *Journal of Philosophy* 111, no. 5 (2014): 229–58.
- Fivush, Robyn. *Family Narratives and the Development of an Autobiographical Self*. Routledge, 2019.
- Fletcher, Guy. "Moral Testimony: Once More with Feeling." In *Oxford Studies in Metaethics*, vol. 11, edited by Russ Shafer-Landau. Oxford University Press, 2016.
- Fricker, Miranda. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press, 2007.
- . "Epistemic Oppression and Epistemic Privilege." *Canadian Journal of Philosophy* 29, supp. 1 (1999): 191–210.
- Fridland, Ellen. "Motor Skill and Moral Virtue." *Royal Institute of Philosophy Supplement* 80 (2017): 139–70.
- Fridland, Ellen, and Matt Stichter. "It Just Feels Right: An Account of Expert

- Intuition." *Synthese* 199, nos. 1–2 (2021): 1327–46.
- Garcia, Manon. *We Are Not Born Submissive: How Patriarchy Shapes Women's Lives*. Princeton University Press, 2021.
- Hills, Alison. "Moral Testimony and Moral Epistemology." *Ethics* 120, no. 1 (2009): 94–127.
- . "Moral Testimony: Transmission Versus Propagation." *Philosophy and Phenomenological Research* 101, no. 2 (2020): 399–414.
- . "Understanding Why." *Nous* 50, no. 4 (2016): 661–88.
- Howell, R.J. "Google Morals, Virtue, and the Asymmetry of Difference." *Nous* 48, no. 3 (2014): 389–415.
- Jacobson, Daniel. "Seeing by Feeling: Virtues, Skills, and Moral Perception." *Ethical Theory and Moral Practice* 8 (2005): 387–409.
- Johnston, Mark. "The Authority of Affect." *Philosophy and Phenomenological Research* 63, no. 1 (2001): 181–214.
- . "How to Speak of the Colors." In *Readings on Color: The Philosophy of Color*, edited by Alex Byrne and David R. Hilbert. Massachusetts Institute of Technology Press, 1997.
- Kelp, Christoph. "Moral Assertion." *Ethical Theory and Moral Practice* 23, nos. 3–4 (2020): 639–49.
- Lewis, Max. "The Norm of Moral Assertion: A Reply to Simion." *Ethical Theory and Moral Practice* 22, no. 4 (2019): 1043–49.
- Lord, Errol. "How to Learn About Aesthetics and Morality Through Acquaintance and Deference." In *Oxford Studies in Metaethics*, vol. 13, edited by Russ Shafer-Landau. Oxford University Press, 2018.
- Malfatti, Federica Isabella. "Can Testimony Transmit Understanding?" *Theoria* 86, no. 1 (2020): 54–72.
- . "On Understanding and Testimony." *Erkenntnis* 86 (2021): 1345–65.
- Margalit, Avishai. *The Ethics of Memory*. Harvard University Press, 2004.
- McDowell, John. "Virtue and Reason." *Monist* 62, no. 3 (1979): 331–50.
- McNaughton, David. *Moral Vision: An Introduction to Ethics*. Oxford University Press, 1988.
- Montague, Michelle. *The Given: Experience and Its Content*. Oxford University Press, 2016.
- Narayan, Uma. "Working Together Across Difference: Some Considerations on Emotions and Political Practice." *Hypatia* 3, no. 2 (1988): 31–48.
- Orwell, George. "A Hanging." In *The Collected Essays, Journalism and Letters of George Orwell*, vol. 1. Penguin, 1970.
- Paul, L. A. "First Personal Modes of Presentation and the Structure of Empathy." *Inquiry* 60, no. 3 (2016): 189–207.
- . "Transformative Choice: Discussion and Replies." *Res Philosophica* 92,

- no. 2 (2015): 473–545.
- . *Transformative Experience*. Oxford University Press, 2014.
- Pettit, Philip. “Realism and Response-Dependence.” *Mind* 100, no. 4 (1991): 587–626.
- Poellner, Peter. *Value in Modernity: The Philosophy of Existential Modernism in Nietzsche, Scheler, Sartre, Musil*. Oxford University Press, 2022.
- Pohlhaus, Gailie. “Relational Knowing and Epistemic Injustice: Toward a Theory of Willful Hermeneutical Ignorance.” *Hypatia* 27, no. 4 (2012): 715–35.
- Schechtman, Marya. *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. Oxford University Press, 2014.
- Simion, Mona. “The Explanation Proffering Norm of Moral Assertion.” *Ethical Theory and Moral Practice* 21 (2018): 477–88.
- Sliwa, Paulina. “Moral Understanding as Knowing Right from Wrong.” *Ethics* 127, no. 3 (2017): 521–52.
- Soteriou, Matthew. *The Mind’s Construction: The Ontology of Mind and Mental Action*. Oxford University Press, 2013.
- Stichter, Matt. “Virtues as Skills, and the Virtues of Self-Regulation.” *Journal of Value Inquiry* 55, no. 2 (2021): 355–69.
- Swartwood, Jason D. “Wisdom as an Expert Skill.” *Ethical Theory and Moral Practice* 16, no. 3 (2013): 511–28.
- Tappolet, Christine. *Emotions, Value, and Agency*. Oxford University Press, 2016.
- Toole, Briana. “Demarginalizing Standpoint Epistemology.” *Episteme* 19, no. 1 (2022): 47–65.
- . “From Standpoint Epistemology to Epistemic Oppression.” *Hypatia* 34, no. 4 (2020): 598–618.
- . “Standpoint Epistemology and Epistemic Peerhood: A Defense of Epistemic Privilege.” *Journal of the American Philosophical Association* 10, no. 3 (2024): 409–26.
- Wiggins, David. “A Sensible Subjectivism?” In *Needs, Values, Truth: Essays in the Philosophy of Value*. Oxford University Press, 1987.
- Wylie, Alison. “Feminist Philosophy of Science: Standpoint Matters.” *Proceedings and Addresses of the American Philosophical Association* 86, no. 2 (2012): 47–76.
- Zagzebski, Linda. “Emotion and Moral Judgement.” *Philosophy and Phenomenological Research* 66, no. 1 (2003): 104–24.

A HEDONIC SUBJECTIVISM

Daniel Pallies

ACCORDING to a standard kind of *subjectivism about well-being*, some things are good for us because of our attitudes towards them. For example, most subjectivists can agree that having a cat is good for me because I love cats. But there is less agreement about what to say if you hate cats. According to one proposal, if you hate cats, then *not* having a cat is good for you. But in recent years, subjectivists have advanced a different proposal: if you hate cats, then having a cat is *bad* for you.¹ The idea is that all theories of well-being, subjectivism included, need to explain not just what is good for us but also what is positively bad for us.² A subjectivist can explain what is good for us by appealing to attitudes like my love of cats and can explain what is bad for us by appealing to attitudes like your hatred of cats.³

Given the need to distinguish between these two kinds of attitudes—those that make their objects good for us and those that make their objects bad for us—we need an account of how the two kinds differ psychologically. What is it about my love of cats that makes it the kind of attitude whose object is good for me? What is it about your hatred of cats that makes it the kind of attitude whose object is bad for you? I argue that the difference between the two kinds of attitudes should be explained in terms of hedonic feelings. If one's attitude towards cat ownership is partly a matter of taking pleasure in having a cat, then having a cat is good for one. If one's attitude towards cat ownership is partly a matter of taking displeasure in having a cat, then one's attitude makes cat ownership bad for one. I call this view *hedonic subjectivism*. The view is hedonic insofar as it gives a central role to pleasure and displeasure in the theory of well-being. But it is not a form of *hedonism*: it does not tell us that only pleasure and displeasure are good or bad for us. Rather, it is a kind of *subjectivism*: it tells us that *the things*

1 Note that these proposals are not mutually exclusive; one could accept both.

2 For arguments in support of the general point that theories of well-being should cover the good and the bad, see Kagan, "An Introduction to Ill-Being." For more specific arguments about how and why subjectivists should cover both the good and the bad, see Ayars, "Attraction, Aversion, and Meaning in Life"; Heathwood, "Ill-Being for Desire Satisfactionists"; and Pallies, "Attraction, Aversion, and Asymmetrical Desires."

3 I apologize if you do not hate cats.

in which we take pleasure or displeasure are good and bad for us, respectively. So if one takes pleasure in cat ownership, then cat ownership itself is good for one, independently of the goodness of the pleasure itself.⁴

I start in section 1 by making the case for distinguishing between a kind of attitude that is essentially connected to well-being and another kind of attitude that is essentially connected to ill-being. Then in section 2, I introduce hedonic subjectivism as an account of the psychological difference between these two kinds of attitudes. In section 3, I offer some preliminary motivations for this view; in section 4, I defend hedonic subjectivism from a battery of objections; and in section 5, I take stock of my conclusions.

1. WELFARE AND ILLFARE ATTITUDES

Many theories of well-being are broadly subjectivist, in the sense of entailing that our attitudes can make a difference to what is nonderivatively good or bad for us. The most obvious example is desire satisfactionism, according to which the objects of only a particular attitude—desire—are nonderivatively good or bad for us.⁵ Desire satisfactionism entails that if I desire to have a cat, then having a cat increases my well-being, and does so in virtue of my having this desire. In other words, cat ownership benefits me in a way that goes beyond the beneficial *effects* of having a cat—the free pest control, for example. So it would not be just as good for me to get these same effects without actually satisfying my desire to have a cat. Going forward, whenever I say that something is good or bad for us, I mean *nonderivatively* good or bad for us, but I leave the qualifier unstated except as an occasional reminder.

Desire satisfactionism is by no means the only theory that tells us our attitudes have this kind of nonderivative significance for well-being. Hybrid

4 In defending hedonic subjectivism, I build upon a recent suggestion by Declan Smithies in “A Hedonic Theory of Desire.” Smithies contends that our desires are relevant to well-being in virtue of their being reducible to pleasure and displeasure. At bottom, then, pleasure and displeasure are the subjective states most fundamentally related to well-being. This leads him to conclude, “We should prefer attitudinal theories of welfare that include all hedonically valenced attitudes, including pleasure as well as desire, within the class of attitudes that determines welfare” (22). I do not take a stand on the nature of desire, but I entirely agree that all pleasant or unpleasant attitudes are directly relevant to well-being. My goal is to develop hedonic subjectivism as the best version of this view.

5 Alternatively, a subjectivist might hold that what is good or bad for us is the *combination* of our attitudes and their objects (wanting-a-cat-and-having-one, for example). This is the so-called *combo view*, as opposed to the *object view*, on which it is only the objects of our attitudes and not the attitudes themselves that are good and bad for us (Lin, “Two Kinds of Desire Theory of Well-Being”). I find the object view more natural and continue to write in a way that assumes it, but the assumption is not doing any heavy lifting here.

theories of well-being allow that our attitudes can make things good or bad for us while retaining a place for more “objective” goods.⁶ And there are versions of the objective list theory on which some objective goods involve our attitudes.⁷ These theories can accommodate the claim that some of our attitudes have nonderivative significance for well-being in the relevant sense. All such views are broadly subjectivist.

Given the truth of some broadly subjectivist view, we can distinguish between those attitudes with *positive* significance and those with *negative* significance: the former make their objects good for us; the latter make their objects bad for us. I call these two kinds of attitudes *welfare attitudes* and *illfare attitudes*, respectively.

Welfare Attitude: A subject bears a welfare attitude towards *p* iff they bear an attitude towards *p* in virtue of which it is good for them that *p*.

Illfare Attitude: A subject bears an illfare attitude towards *p* iff they bear an attitude towards *p* in virtue of which it is bad for them that *p*.

I argue that welfare and illfare attitudes are reducible to pleasure and displeasure, respectively.⁸

First, it needs to be shown that welfare and illfare attitudes really are distinct *kinds* of attitude, as opposed to being different ways of describing the same kind of attitude. The latter view is exemplified by a simple version of desire satisfactionism.

Desire Theory of Welfare Attitudes: A subject bears a welfare attitude with strength *s* towards *p* iff they *desire* that *p* with strength *s*.

Desire Theory of Illfare Attitudes: A subject has an illfare attitude of strength *s* towards *p* iff they *desire* that $\sim p$ with strength *s*.

6 Lovett and Riedener, “The Good Life as the Life in Touch with the Good”; and Wall and Sobel, “A Robust Hybrid Theory of Well-Being.”

7 Fletcher, “A Fresh Start for the Objective-List Theory of Well-Being”; and Murphy, *Natural Law and Practical Rationality*.

8 Why use the terms ‘welfare attitude’ and ‘illfare attitude’, as opposed to the more natural ‘positive attitude’ and ‘negative attitude’? The problem with the latter terminology is that ‘positive attitude’ and ‘negative attitude’ have psychological meanings in natural language. I do not want to assume that to have an attitude towards something in virtue of which it is good for us, it is necessary or sufficient to bear a *positive attitude* towards it, where ‘positive attitude’ is used in the ordinary psychological sense.

On this view, the distinction between welfare and illfare attitudes comes down to a distinction between two ways in which a single psychological attitude—desire—is relevant to well-being.

The way to see that this kind of view is mistaken is to notice that welfare and illfare attitudes can diverge: one can have a stronger welfare attitude towards p than one's illfare attitude towards $\sim p$, and vice versa.⁹ So it cannot be that welfare attitudes and illfare attitudes are really the same attitude described in different ways.

To see how and why welfare and illfare attitudes can diverge, we can start with the following case.

Three Performers: Three violinists, Newbie, Diva, and Paycheck, are invited to perform at a party. Newbie is earnest and grateful; she is excited simply to have been invited. Diva is grumpy and arrogant; she considers this sort of work beneath her talents. Paycheck just wants her paycheck. Newbie does not expect to have a captive audience, though she delights at the idea of being the center of attention. Diva does expect to have a captive audience, though she hates the idea of *not* being the center of attention. Paycheck has no expectations and does not care much either way. As it happens, the three musicians are so engrossed in their performance that they do not notice and never learn whether the partygoers pay any attention.

Newbie and Diva share a preference to have the partygoers' attention, but their attitudes nevertheless differ in ways that are relevant to what is good or bad for them. To see this, we can start by comparing their levels of well-being on the assumption that the partygoers are in fact paying attention to the performance. If the partygoers are paying attention, then that is very good for Newbie. Having a captive audience is a dream come true for her; it is the culmination of her ambition as an aspiring performer. By contrast, it is not at all a dream come true for Paycheck or Diva. Paycheck does not care whether she has a captive audience, and Diva more or less takes it for granted that everyone pays attention to her. Neither Paycheck nor Diva would feel gratified or appreciative if they were to learn that the partygoers are enthralled. So while it is plausible that having the partygoers' attention is very good for Newbie, the same cannot be said for Diva or Paycheck.

9 For more detailed defenses of this claim, see, for example, Kelley, "Well-Being and Alienation"; Ayars, "Attraction, Aversion, and Meaning in Life"; Heathwood, "Ill-Being for Desire Satisfactionists"; Mathison, "Asymmetries and Ill-Being"; and Pallies, "Attraction, Aversion, and Asymmetrical Desires."

Now consider the possibility that the audience members are not paying attention to the performance. If the partygoers are not paying attention, then that is very bad for Diva. She hates the idea of her music being ignored or overlooked; she considers it a total tragedy to be ignored while she is performing. By contrast, being ignored is not a tragedy for Paycheck or Newbie. Paycheck does not care whether she is ignored, and Newbie more or less takes it for granted that she will be ignored. Neither Paycheck nor Newbie would feel at all upset or insulted if they were to learn that the partygoers are not enthralled by the performance. So while it is plausible that lacking the partygoers' attention is very bad for Diva, the same cannot be said for Newbie or Paycheck.

Putting all these comparisons together, we arrive at the following set of judgments:

	Significance for Newbie's well-being	Significance for Diva's well-being	Significance for Paycheck's well-being
Captive audience	High positive significance	Little or no significance	Little or no significance
No captive audience	Little or no significance	High negative significance	Little or no significance

This pattern in the violinists' levels of well-being can be easily explained if we allow that welfare and illfare attitudes can diverge. Newbie bears a strong welfare attitude towards having a captive audience but lacks a strong illfare attitude towards lacking a captive audience. The reverse is true for Diva. Thus, welfare and illfare attitudes must be different kinds of attitudes; they are not merely desires or preferences described in different ways.

Given that welfare and illfare attitudes differ in kind, we need a theory of the difference. The case of the three violinists is helpful insofar as it provides some initial *paradigms* of welfare and illfare attitudes, but it would be better to have a general theory. This is what hedonic subjectivism promises.

2. HEDONIC SUBJECTIVISM

On the view I defend, all welfare attitudes involve pleasure, and illfare attitudes involve displeasure. The theory is simple:

Welfare Attitudes to Pleasure: A subject bears a welfare attitude towards *p* iff they are disposed to take pleasure in *p*.

Illfare Attitudes to Displeasure: A subject bears an illfare attitude towards *p* iff they are disposed to take displeasure in *p*.

Recall that by definition, welfare and illfare attitudes are all and only those attitudes that make their objects nonderivatively good and bad for us, respectively. So the hedonic view has the following implications for well-being.

Pleasure-Welfare: It is good for a subject that *p*, in virtue of that subject's attitude towards *p*, iff that subject is disposed to take pleasure in *p*.

Displeasure-Welfare: It is bad for a subject that *p*, in virtue of that subject's attitude towards *p*, iff that subject is disposed to take displeasure in *p*.

To see the general idea here, consider the Three Performers case. Newbie's attitude towards being the center of attention can be described in various ways: she loves the idea, she regards it as a dream come true, and so on. According to hedonic subjectivism, what matters most fundamentally from the standpoint of well-being is that in having these attitudes, she *takes pleasure* in being the center of attention. That is why being the center of attention is good for her. The same goes for Diva. She hates the idea of being ignored, she regards it as a tragedy, and so on, but what matters most fundamentally from the standpoint of well-being is that these are ways of *taking displeasure* in being ignored. That is why it is bad for her to be ignored.

That is the gist of hedonic subjectivism. But there are a number of details to be clear about before we can start to assess the view.

2.1. *Pleasure and Displeasure*

By 'pleasure' and 'displeasure', I mean nothing more than *pleasant experience* and *unpleasant experience*, respectively.¹⁰ Some philosophers and scientists suggest that there can be pleasures and displeasures that are not experiences; if so, they are not my concern here.¹¹ The nature of pleasant and unpleasant experience is a large subject, but paradigm instances are sensory experiences like the pleasure of smelling baking bread or the displeasure of smelling garbage. In addition to these straightforwardly sensory cases, there are also more intellectual and emotional pleasures and displeasures—the pleasures of daydreaming about fame and fortune, of meeting important goals, and of receiving compliments, as well as the displeasures of ruminating about possible disasters, of failing to meet one's goals, and of being insulted. Many of the pleasures and displeasures most clearly implicated in welfare and illfare attitudes are cognitive or emotional.

10 Here and elsewhere, I prefer the term 'displeasure' to the less inclusive 'pain'. Displeasure covers many unpleasant experiences that are not painful (for example, itchiness, vertigo, and nausea).

11 See, for example, Berridge and Winkielman, "What Is an Unconscious Emotion?"; and Peciña et al., "Hedonic Hot Spots in the Brain."

2.2. *Taking Pleasure or Displeasure in Something*

When we say that someone *takes pleasure in p*, we are not merely saying that *p* caused them to feel pleasure. It may be that due to some strange causal chain, solar flares cause me to feel pleasure. Even so, it would not follow that I take pleasure in solar flares. A better proposal is that to take pleasure in something is to have a pleasant experience that *represents* that something.¹² So to take pleasure in the fact that a solar flare is occurring, I would have to have a pleasant experience *as of* a solar flare occurring.

Since our experiences can represent things that do not exist, it follows that we can take pleasure in things that do not exist. For example, suppose I am daydreaming about having a flying car. My imaginative experience is pleasant, and the experience represents my having a flying car. It follows that I am taking pleasure in *having a flying car*. But of course I do not have a flying car. In this sort of case, we might ordinarily say that I take pleasure in *the thought* of having a flying car, thereby avoiding the implication that the car really exists. I sometimes use this expression when it is more natural to do so, but I also stipulate that as I understand the “taking pleasure” relation, there is no implication that our pleasures represent things that really exist.¹³ The important point is that if a subject is disposed to have pleasant experiences as of *p*, then *p* is good for them. All the same considerations apply equally to displeasure, so if a subject is disposed to have unpleasant experiences as of *p*, then *p* is bad for them.

2.3. *Intrinsic and Instrumental Attitudes*

If I desire something, we can ask whether I desire it *instrumentally* (as a means to an end) or *intrinsically* (as an end in itself). Among proponents of desire-based subjectivism, it is typical to claim that the objects of only our *intrinsic* desires are nonderivatively good for us. It is not clear that hedonic subjectivists can make the same claim because it is not clear that it makes sense to ask whether I take pleasure in something *instrumentally* or *intrinsically*.¹⁴

In any case, I do not think that hedonic subjectivists need to avail themselves of this distinction. Proponents of desire-based subjectivism need the distinction—or typically take themselves to need the distinction—because they hold that the objects of only our intrinsic desires are good for us.¹⁵ This restriction is motivated by the following sort of case: I desire a ticket to the carnival, purely as a means to attending the carnival, and although I do get a ticket,

12 Smithies uses a similar case to argue for a similar point (“A Hedonic Theory of Desire,” 18).

13 See also Feldman, “Two Questions About Pleasure,” 72.

14 Cf. Feldman, *Pleasure and the Good Life*, 57–58.

15 For a counterexample to this general trend, see Heathwood, “Desire-Fulfillment Theory,” 139.

I am nevertheless prevented from attending the carnival. It is easy to get the intuition that I did not benefit at all—I got nothing I really wanted. Proponents of desire-based subjectivism tell us that I did not benefit because the objects of only intrinsic desires are nonderivatively good for us.

The case of the carnival ticket should not motivate hedonic subjectivists to restrict their theory in the same way. For the hedonic subjectivist, the question is whether I take pleasure in having a ticket. If I do, then I benefit from having a ticket; otherwise, I do not. And these predictions seem to be correct. Suppose I take pleasure in the simple fact of having a ticket to the carnival: thinking about or imagining having a ticket is itself pleasant, quite independently of thinking about or imagining going to the carnival. Then it is plausible that I benefit from having a ticket, even if I do not end up going to the carnival. If, on the other hand, I merely take pleasure in the thought of attending the carnival and take no pleasure in having a ticket, then I do not seem to benefit from having a ticket. So hedonic subjectivism gets the right results without needing to appeal to a distinction between intrinsic and instrumental versions of the relevant attitude.

2.4. *Instrumental Versus Noninstrumental Value*

Turning from instrumental attitudes to instrumental value, note that hedonic subjectivism is not a view about what is instrumentally valuable. It is easy to misread hedonic subjectivism as the view that some things are instrumentally good for us because we take pleasure in them, and some things are instrumentally bad for us because we take displeasure in them. But in fact, the view tells us that the objects of pleasure and displeasure are noninstrumentally good and bad for us, respectively, *because* we take pleasure or displeasure in them. Thus, when Newbie takes pleasure in the thought of being the center of attention, *being the center of attention* is noninstrumentally good for her. If she is the center of attention, then her life includes one more welfare good than it would if she were not the center of attention, even if she would feel the same amount of pleasure either way.

This point is particularly worth noting because it marks the central difference between hedonic subjectivism and hedonism, including versions of hedonism that attempt to accommodate the thought that, for example, it is better for Newbie to be the center of attention even if she feels the same amount of pleasure either way. For example, Fred Feldman describes a view on which pleasure is the only thing that is good for us, but pleasures directed at “true objects” are better than those directed towards “false objects.”¹⁶ This *truth-adjusted hedonism* allows us to say that it is better for Newbie to be the center of attention since in that case, her pleasure has a “true object” and is therefore better for

16 Feldman, *Pleasure and the Good Life*, 111.

her. But truth-adjusted hedonism does not allow us to say what nonhedonists typically want to say about cases like this—namely, that something *other than pleasure* (in this case, being the center of attention) is noninstrumentally good for the subject. There are other important differences between Feldman's hedonism and hedonic subjectivism, but this is the central difference.¹⁷ Feldman's hedonism is still a version of *hedonism*. Hedonic subjectivism is not, because it affirms that things other than pleasures are noninstrumentally good for us.¹⁸

2.5. *Subjective and Objective Well-Being*

Hedonic subjectivism tells us that the objects of our pleasant and unpleasant attitudes are good and bad for us, respectively, but it does not tell us that *only* the objects of these attitudes are good or bad for us. In other words, hedonic subjectivism is not a complete theory of well-being. Rather, it is a theory about the subjective component of well-being; it describes how and why our subjective attitudes make their objects good or bad for us. I leave open the possibility that there may also be an objective component to well-being; it may be that some things are good or bad for us independently of our attitudes towards them. Perhaps knowledge and achievement are objectively good, whereas ignorance and failure are objectively bad.

This leaves open an important question regarding the goodness and badness of pleasure and displeasure themselves. Like most philosophers of well-being, I am confident that pleasure itself is nonderivatively good for us and that

17 For example, Feldman's attitudinal hedonism tells us that "you can take pleasure in something at a time when you don't *feel* any pleasure" (*Pleasure and the Good Life*, 56, emphasis original). And the truth-adjusted version of the view tells us that all else being equal, pleasures taken in true objects are ten times better than pleasures taken in false objects (112–13).

18 One might hold that the only important differences between theories of well-being are differences in their implications regarding subjects' levels of well-being. If so, then there might not be any important difference between hedonic subjectivism and truth-adjusted hedonism. For it could be that "true pleasures" are noninstrumentally better than "false pleasures" (according to truth-adjusted hedonism) to exactly the same degree that the objects of our pleasures are good for us (according to hedonic subjectivism). Then the two views would always issue the same verdicts regarding subjects' levels of well-being. But it would be a mistake to conclude that there is no important difference between them. To see why, consider that Feldman could go further: he could hold that although pleasure is all that is good for us, its degree of goodness is determined by the balance of knowledge, achievement, and friendship in one's life. By design, this view might never diverge from a certain kind objective list theory with respect to its verdicts about subjects' levels of well-being. But objective list theorists nevertheless prefer their own view. They insist that pleasure is not all that is good for us and that other goods do not make a difference to how good our pleasures are for us. Similarly, subjectivists may insist on the same points: pleasure is not all that is good for us, and other goods do not make a difference to how good our pleasures are for us.

displeasure itself is nonderivatively bad for us. I am less confident about whether their goodness and badness belong to the subjective side of well-being or to the objective side. In support of the view that they belong to the subjective side, I am inclined to think that all pleasures and displeasures represent themselves in some sense—either because all experiences involve a kind of primitive self-awareness or because pleasure and displeasure specifically are directed towards themselves.¹⁹ If so, it turns out that whenever we experience pleasure or displeasure, we count as taking pleasure or displeasure in the experience itself. Hedonic subjectivism could then capture the goodness and badness of pleasure and displeasure as part of the subjective side of well-being. But I admit that this proposal is highly speculative, and I cannot defend it in any detail here. Suffice it to say that if I cannot capture the goodness and badness of pleasure and displeasure with hedonic subjectivism, then I would accept that their goodness and badness belong to the objective side of well-being. Either way, I accept that pleasure and displeasure are nonderivatively good and bad for us, respectively.

3. PRELIMINARY MOTIVATIONS

The preliminary case for hedonic subjectivism is that it builds upon the appealing features of both hedonism and traditional forms of subjectivism. Starting with traditional forms of subjectivism, we have already seen one way in which hedonic subjectivism builds upon this view. All subjectivists need to distinguish between welfare and illfare attitudes, and hedonic subjectivism allows them to do so in a straightforward way.

Hedonic subjectivism also fleshes out a claim that a number of subjectivists have defended in recent years—namely, that the attitudes relevant to well-being must be *affective*.²⁰ The thought is that if one merely desires something in the sense that one is motivated to bring it about, then the satisfaction of the desire is not necessarily good for one, nor is its frustration necessarily bad for one. For example, suppose I am motivated to visit a depressed friend out of a sense of duty, or suppose I have a strange disposition to turn on radios in my vicinity.²¹ In both cases, I may or may not successfully do what I am motivated to do, but either way, there does not appear to be any immediate

19 For defenses of the view that experience essentially involves self-awareness, see Kriegel, *Subjective Consciousness*; and Strawson, “Self-Intimation.” For defenses of the view that affective experiences specifically are self-directed, see Barlassina and Hayward, “More of Me!”; and Pallies, “An Honest Look at Hybrid Theories of Pleasure.”

20 Fanciullo, “Alienation, Engagement, and Welfare”; and Heathwood, “Which Desires Are Relevant to Well-Being?”

21 Quinn, “Putting Rationality in Its Place,” 32; and Gosling, *Pleasure and Desire*, 86.

impact on my well-being. Some subjectivists argue that this is because I am not *affectively engaged* in my activity. And what is it to be affectively engaged? Here, subjectivists tend to rely on lists of examples of affective engagement or on stereotypical features of affective engagement.²² But if the relevant notion is to do so much heavy lifting in the theory of well-being—distinguishing those attitudes that are directly relevant to well-being from those that are not—then it would be better to have a more contentful description. Hedonic subjectivism fills the lacuna: being affectively engaged with something, in the relevant sense, is nothing more than taking pleasure or displeasure in it.

Hedonic subjectivism also builds on the appeal of hedonism. Part of the appeal of hedonism is its explanation of how and why things are good or bad for us. Hedonists tell us that whenever anything is good or bad for us, its goodness or badness derives from the goodness or badness of pleasure or displeasure. This explanation captures a heterogeneous assortment of goods and bads—pasta dinners, walks in the park, broken bones, bad breakups, etc.—and it does so in an appealingly simple and unified way. Moreover, the explanation has an appealingly solid foundation because the well-being significance of pleasure and displeasure is nigh indubitable. This is presumably why pleasure and displeasure are the *only* candidate goods and bads whose well-being significance is recognized by almost all philosophers of well-being. All told, then, it seems to me that hedonists have a highly appealing explanation of why various things are good or bad for us. The explanation is simple and unified, it captures many heterogeneous goods and bads, and it has an epistemically secure foundation.

Hedonic subjectivism shares in these virtues. Like hedonists, a hedonic subjectivist explains why things are good or bad for us by appealing to the ways in which those things are related to our experiences of pleasure or displeasure. There are differences: while a hedonist says that things are (derivatively) good or bad for us because they *cause* pleasure or displeasure, a hedonic subjectivist says that things are (nonderivatively) good or bad for us because we *take pleasure or displeasure in them*. But the hedonic subjectivist explanation has many of the same virtues as the hedonist explanation. It is simple and unified, it captures many heterogeneous goods and bads, and it has an epistemically secure foundation. Hardly anyone doubts that pleasure and displeasure have important roles in well-being. Hedonic subjectivism merely gives them more expansive roles.

At the same time, hedonic subjectivism also avoids hedonism's most serious problem: the experience machine objection. Hedonism implies that it is

22 For lists of examples, see Chang, "Can Desires Provide Reasons for Action?" 80–81; and Fanciullo, "Alienation, Engagement, and Welfare," 16–17. For descriptions of stereotypical features, see Chang, "Can Desires Provide Reasons for Action?" 68–69; and Heathwood, "Which Desires Are Relevant to Well-Being?" 674–75.

no better to lead a rich and fulfilling life than it is to have all the experiences as of leading a rich and fulfilling life.²³ My sense is that this is the most damaging implication of hedonism—the implication that most often motivates nonhedonists in their rejection of hedonism. But hedonic subjectivism does not share this damaging implication because it allows that a rich and fulfilling life includes many welfare goods that are absent from life in the experience machine. In a rich and fulfilling life outside the experience machine, one takes pleasure in many things; the objects of those pleasures really exist, and they are good for one. In the experience machine, the objects of one's pleasures do not exist, so one's life is missing many goods that are present in a life outside the machine.

All this leads me to think that hedonic subjectivism is worth investigating. But ultimately, the true test of a theory is whether it can withstand the objections leveled against it. With that in mind, the goal of the next section is to answer the objections.

4. OBJECTIONS TO HEDONIC SUBJECTIVISM

4.1. *Objections to Subjectivism*

Some general objections to subjectivism can be modified to target hedonic subjectivism specifically. For example, consider John Rawls's case of the person who strongly desires to do nothing but count grass in Harvard Yard.²⁴ Or consider Derek Parfit's case of meeting a stranger on a train and desiring years later that the stranger is in good health.²⁵ Finally, consider Bernard Williams's case in which a man named Sam desires to drink the glass of clear liquid on the table because he believes that it is gin when it is in fact petrol.²⁶ There are many other objections in this vein, alleging that a subject's desire is *too trivial*, *too remote*, or *too ill-informed* to count towards the subject's well-being. The objections can be easily modified to target hedonic subjectivism. We can imagine that the grass counter takes pleasure in counting grass, that you take pleasure in the thought of the distant stranger's health, and that Sam takes pleasure in the thought of drinking the glass of clear liquid. The point of the objections remains the same—the objector alleges that the desires are too trivial, too remote, or too ill-informed to count towards the subject's well-being.

23 Lin, "How to Use the Experience Machine"; and Nozick, *Anarchy, State, and Utopia*, 73–74.

24 Rawls, *A Theory of Justice*, 432.

25 Parfit, Derek, *Reasons and Persons*, 151.

26 Williams, "Internal and External Reasons."

I want to set these objections aside. They are important objections to subjectivism, but they have nothing to do with hedonic subjectivism specifically. In the face of the objections, a hedonic subjectivist's options are the same as those of any other subjectivist. A hedonic subjectivist might try to explain away our intuition that the subjects do not benefit from satisfying their desires.²⁷ Or they might concede that the subjects do not benefit, and modify their subjectivism so as to explicitly exclude trivial, distant, or ill-informed desires. I think that subjectivists should do without these modifications, but this is not the place to engage in that particular debate. The important point is that a hedonic subjectivist could modify their theory in the same ways.²⁸ They could say that we do not benefit from the objects of our pleasures when those objects are trivial, as in the case of grass counting. They could say that we do not benefit from taking pleasure in something for someone else's sake, as in the case of the distant stranger. And they could say that we do not benefit from the objects of our pleasures when we are ill-informed about those objects, as in the case of the glass of petrol.

More important for my purposes are objections that do target hedonic subjectivism specifically—namely, objections alleging that some welfare attitudes do not involve pleasure, or some illfare attitudes do not involve displeasure. It is to these arguments I now turn.

4.2. *Desire*

We can start with a straightforward argument that a desire-based subjectivist might give against hedonic subjectivism.

Desire Objection:

- P1. Some desires do not involve pleasure but are welfare attitudes.
- P2. If P1, then hedonic subjectivism is false.
- C. Hedonic subjectivism is false.

As straightforward as this argument seems, I believe it is actually something of a red herring. The argument cannot be evaluated on its own; it needs to be paired with a theory of desire. There are, after all, many different theories of

27 For a useful catalog of various attempts along these lines, see Heathwood, "Desire-Fulfillment Theory."

28 A reviewer worries that this move is unavailable to hedonic subjectivists. While it is obvious that I can desire something for someone else's sake, it is less obvious that I can take pleasure in something for someone else's sake. In response, I agree the latter expression is extremely awkward, but even so, I am inclined to think we can indeed take pleasure in something for another's sake. After all, it is natural to say that I am pleased for your sake that you were promoted. And I am inclined to think that this is merely a less awkward way of saying that *for your sake*, I take pleasure in the fact that you were promoted.

desire.²⁹ And once we have a theory of desire at hand, we can use it to formulate an argument that makes no appeal to desire. To illustrate, suppose the objector tells us that they have in mind an *evaluative* theory of desire: they think that desires are evaluative beliefs, and at least some of these desires are welfare attitudes despite not involving pleasure. Now we can evaluate the argument, but we can also see there was no need to state the argument in terms of desire in the first place. The objector could have simply stated the objection in terms of evaluative beliefs. After all, if some evaluative beliefs are welfare attitudes despite not involving pleasure, then hedonic subjectivism is false, whether or not these evaluative beliefs are *desires*.

With that in mind, I set aside questions about the nature of desire. I instead deal directly with what I take to be the most plausible proposals for welfare attitudes that do not essentially involve pleasure, and illfare attitudes that do not essentially involve displeasure. I argue that these proposals fail, but I make no claims about the nature of desire.

4.3. Evaluations

Start with the proposal that positive evaluations are welfare attitudes. This proposal comes in a number of different varieties. One might hold that a subject bears a welfare attitude towards some object if they *believe that the object is good for them* or if they *value* that object or if they *see it as noncomparatively good*. To streamline the discussion, I say that a subject *comprehensively endorses* some object just in case they positively evaluate it in *all* of these ways. And I assume that we can comprehensively endorse some object without taking pleasure in it.³⁰ The question then is whether the objects of our comprehensive endorsements are good for us in virtue of their being positively evaluated in these ways. If they are, then hedonic subjectivism is false. I contend that they are not.

29 Theories of desire are highly heterogenous. Some hold that desires are representations of goodness. See, for example, Gregory, *Desire as Belief*; Oddie, *Value, Reality, and Desire*, ch. 3; and Stampe, "The Authority of Desire," 359–62. Others hold that they are functional states. See Lewis, "Psychophysical and Theoretical Identifications"; Millikan, *Language*, 171–73; and Papineau, "Representation and Explanation," 562–65. Others understand desires in terms of attention. See Scanlon, *What We Owe to Each Other*, 38–42; and Schroeder, *Slaves of the Passions*, ch. 8. Still others understand them in terms of learning. See Schroeder, *Three Faces of Desire*. Yet others in terms of hedonic feelings. See Smithies, "A Hedonic Theory of Desire."

30 For evaluative belief, see Dorsey, "Subjectivism Without Desire." For valuing, see Raibley, "Values, Agency, and Welfare"; and Tiberius, *Well-Being as Value Fulfillment*. For seeing as noncomparatively good, see Ayars, "Attraction, Aversion, and Meaning in Life."

To see this, we need to imagine a case in which someone comprehensively endorses something without being at all disposed to take pleasure in it.³¹ To avoid a possible source of confusion, we should imagine that the object of the comprehensive endorsement is not the sort of thing that might be good for them independently of their attitudes towards it. (We should not imagine that the subject comprehensively endorses knowledge or achievement, for example.) So if it is nevertheless good for the subject, it is good for them in virtue of their attitudes towards it.

With that in mind, consider the case of Eva. When Eva walks to work, she has a choice of whether to walk through the city or walk through the park. Her parents always told her that walking through the park is better than walking through the city, and as a child, she unreflectively accepted their claims. Now, as an adult, Eva comprehensively endorses her walks through the park: she believes they are good, she values them, and she sees them as good for her. But she takes no pleasure in these walks. Indeed, she feels no pleasure in connection with the park at all. She does not enjoy the sights or sounds of the park, nor does she feel pleasant emotions in connection with the park—she does not look forward to her walks with pleasure, nor does she feel pleasantly gratified about having completed them. Although she comprehensively endorses her walks through the park, these evaluations are untouched by any pleasant feelings.

Eva's case is unusual. Usually, if someone comprehensively endorses some activity, they also take some pleasure in the activity. If a painter comprehensively endorses painting, then she probably takes some pleasure in painting, and if she ceases to endorse it, then she will also cease to take pleasure in it. This is a real loss; it is the sort of loss that is familiar from cases of depression. But no such loss would occur if Eva were to stop comprehensively endorsing her walks through the park. Suppose that the process is gradual, and by the time she loses those attitudes entirely, she does not remember ever having them. She does not enjoy her walks through the park, but she never enjoyed them; she merely saw them as good—and good for her—and so on. It does not seem to me that there is any loss of well-being here.

I conclude that positive evaluations, in the absence of pleasure, are not welfare attitudes. It is true that the objects of these evaluations may be good for

31 In principle, one could claim that it is not possible to evaluate something in any of these ways unless one is disposed to take pleasure in it. Valerie Tiberius comes closest to making this claim: she suggests that if someone is not emotionally invested in what they claim to value, then we have cause to be skeptical that they value it (*Well-Being as Value Fulfillment*, 59). On the other hand, Dale Dorsey ("Subjectivism Without Desire," 412), Jason Raibley ("Values, Agency, and Welfare"), and Alisabeth Ayars ("Attraction, Aversion, and Meaning in Life") are all clear that relevant evaluations need not involve pleasure.

us in an attitude-independent way. It is also true that we often take pleasure in the thought of things that we regard as good (or good for us). But merely regarding something as good (or good for us) does not make it the case that it is good for us.

4.4. *Mixed Hedonic Feelings*

Among philosophers who distinguish between attraction and aversion, some have claimed that *both* attraction and aversion are associated with *both* pleasure and displeasure. They say that attraction is associated with the unpleasant *disappointment* that we feel when we are attracted to some prospect and discover that it has not come to pass, and aversion is associated with the pleasant *relief* that we feel when we are averse to some prospect, and we discover that it has not come to pass.³² Relatedly, Declan Smithies and Jeremy Weiss claim that attraction is distinct from pleasure on the grounds that “feelings of attraction themselves are not always pleasurable; for example, bodily cravings and feelings of unrequited love can be intensely painful.”³³

If attraction is a welfare attitude and if it is possible to be attracted to something in virtue of one’s *unpleasant* feelings towards it, then hedonic subjectivism is false. Similarly, if aversion is a welfare attitude and if it is possible to be averse to something in virtue of one’s *pleasant* feelings towards it, then hedonic subjectivism is false. So these are the possibilities that hedonic subjectivists must reject.

To make things concrete, imagine that Romeo is pining away for Juliet. His yearning is painful; he finds it painful to be without her. He would be unpleasantly disappointed if he were to learn that they will not be together and would be pleasantly relieved if he learned that they will. Clearly, Romeo bears welfare attitudes towards the prospect of being with Juliet—it would be very good for him to be with her, in virtue of his attitudes towards her. The first question to ask is whether Romeo’s welfare attitudes are partly a matter of his having the *unpleasant* experiences. If his welfare attitudes are reducible to displeasure in this way, then hedonic subjectivism is false. But fortunately for hedonic subjectivists, there is a better interpretation of the case. We should say that Romeo has both welfare *and* illfare attitudes: he has welfare attitudes towards the thought of being with Juliet and illfare attitudes towards the thought of *not* being with Juliet. The pleasant relief he would feel at being together with Juliet is a welfare attitude towards being with her, but his unpleasant experiences—his

32 Ayars, “Attraction, Aversion, and Meaning in Life,” 20; Schroeder, *Three Faces of Desire*, 132; and Sinhababu, *Humean Nature*, 48.

33 Smithies and Weiss, “Affective Experience, Desire, and Reasons for Action,” 44.

unpleasant pangs of longing and his unpleasant disappointment—are instances of illfare attitudes, not welfare attitudes. Those attitudes make it the case that being *without* Juliet is nonderivatively *bad* for him.

It is also unlikely that Romeo's pangs of longing are purely unpleasant. Lovestruck longing, like many desires, is accompanied by both pleasure and displeasure. Sometimes one has pleasant thoughts ("It would be so nice to be with them again!"), and sometimes one has unpleasant thoughts ("It's so hard to be without them!"). So while it can be unpleasant for Romeo to think about Juliet, it is not *uniformly* unpleasant—it is unpleasant insofar as he thinks about her absence, and pleasant insofar as he thinks about being with her again. The experience is paradigmatically *bittersweet*, and bittersweet emotions are not hedonically neutral but rather hedonically ambivalent. So hedonic subjectivism delivers a plausible result: it is good for Romeo to be with Juliet and bad for him to be without Juliet.

Roughly the same story is true of bodily cravings such as hunger and thirst. Suppose you are sitting down to eat at a restaurant, and you are ravenously hungry. You see someone else being served a delicious meal, and you have a strong craving to get the same meal yourself. Under such conditions, you are likely to oscillate between unpleasant thoughts ("I'm so hungry!") and pleasant thoughts ("The food is going to be so good!"). Again, hedonic subjectivism delivers a plausible result: it is good for you to get the delicious meal and bad for you to be without it. Cases of mixed hedonic feelings do not pose a problem for hedonic subjectivism.

4.5. Pure "Wanting"

An opponent of hedonic subjectivism might argue that the view is undermined by results in modern neuroscience, particularly work on "wanting" and "liking." Kent Berridge argues that "wanting" and "liking" are distinct modules in the brain: very broadly, "wanting" motivates pursuit of reward, whereas "liking" is associated with enjoyment of a reward once attained. And in experimental settings, these modules can be dissociated from one another so that a test subject doggedly pursues a reward despite failing to appreciate it once it has been attained, or fails to pursue a reward despite appreciating it greatly once it is attained.³⁴ These experimental results might move us to draw some quick conclusions: "wanting" is a welfare attitude; "liking" is pleasure; and since "wanting" and "liking" are doubly dissociable, there are welfare attitudes that do not involve pleasure.³⁵

34 Berridge, "Wanting and Liking"; and Kringelbach and Kent, "The Joyful Mind."

35 Thanks to an anonymous reviewer for urging me to address this point.

In fact, however, it is far from obvious that “liking” is pleasure or that “wanting” is a propositional attitude at all. Berridge himself rejects the tempting identification of “wanting” with *desire* in the ordinary sense. He writes:

Incentive salience [another term for “wanting”] as a module is only one type of wanting. It is not the one we are most aware of in daily life nor the type of desire that has been the greatest focus of philosophers.³⁶

And he speculates:

Perhaps a reason for the difference is that incentive salience is mediated chiefly by subcortical brain mechanisms, whereas cognitive forms of desire are more dependent on higher cortex-based brain systems.³⁷

My own view is that “wanting” and “liking” are best understood as *subpersonal* modules whose relations to the more familiar, person-level categories of desire and pleasure are uncertain.³⁸ In any case, it is not necessary to get into the weeds here. For my purposes, it is enough to argue that “wanting” by itself is insufficient for having a welfare attitude. So whether or not “wanting” is desire or any other kind of propositional attitude, there is no problem here for hedonic subjectivism.

To see why “wanting” is not a welfare attitude, it is helpful to note that part of what the distinction between “wanting” and “liking” is supposed to explain is a certain prototypical pattern of behavior in drug addicts: an addict is increasingly motivated to seek out the drug as they become more and more sensitized to it, but the pleasure they get from the drug does not increase accordingly. According to the *incentive-sensitization theory* of addiction, this pattern is the result of a breakdown in the normal cooperation in the “wanting” and “liking” systems: the drug causes increased “wanting” even as “liking” remains constant or decreases. The implication is that if we want to get a sense of what pure “wanting” is like for the subject, then we can imagine the cravings or urges of a drug addict who takes no pleasure in the thought of getting the drug. The question is whether it is noninstrumentally good for this addict to get the drug. Of course, it may be *better* for them to get the drug than to go without the drug (at least in the short term). In all likelihood, they will feel unpleasantly frustrated and anxious about lacking the drug, so it can be noninstrumentally bad for

36 Berridge, “Wanting and Liking,” 379 (clarification added).

37 Berridge, “Wanting and Liking,” 379.

38 Berridge seems more sanguine about the identification of pleasure with “liking.” But since he contends that liking need not be experienced, it is not clear that this is the same sense of pleasure that is most important to philosophers. See Berridge and Winkielman, “What Is an Unconscious Emotion?”

them to lack the drug. There may also be a noninstrumental good associated with taking the drug—namely, the pleasure it causes. All of this can explain why the addict is better-off getting rather than lacking the drug (at least in the short term). But none of it is germane to the question of whether *taking the drug* is noninstrumentally good for the addict.

Imagine a nearby variant of a case considered by Parfit.³⁹ A drug addict is given an unlimited supply of their drug of choice, so they will never feel unpleasantly frustrated or anxious about lacking it. And for whatever reason, they do not get any pleasure from the drug—they get no pleasure from taking it, and they take no pleasure in thinking about it. They simply have an urge or craving to take the drug, recurring at regular intervals, that quickly goes away when they take it. Perhaps they have a device on their wrist that allows them to diffuse the drug into their bloodstream at the press of a button, so they can immediately act on the urges and cravings they feel. Even so, it is not at all plausible that this addict benefits from regularly satisfying their urges and cravings.⁴⁰ They are not better-off than an otherwise similar person who lacks the urges, cravings, and drugs. So I conclude that pure “wanting” is not a welfare attitude.

4.6. *Philosophical Vulcans*

David Chalmers employs an argument that threatens hedonic subjectivism in an indirect way. The objection does not tell us that some *particular* attitude is not pleasant, despite being a welfare attitude. Nor does it tell us that some *particular* attitude is not unpleasant, despite being an illfare attitude. Instead, the argument alleges that there must be *some or other* attitudes that meet these descriptions, though the argument does not tell us much about them.

Chalmers’s argument appeals to creatures that he calls *philosophical Vulcans*. Unlike philosophical zombies, Vulcans do have phenomenal experience. But they are entirely devoid of *hedonic experience*; they are incapable of having any sort of pleasant or unpleasant experiences at all. Chalmers tells us:

39 Parfit, *Reasons and Persons*, 479.

40 In a recent discussion of objections to desire satisfactionism, Chris Heathwood offers a related treatment of the case. He contends that taking the drug is in fact good for the addict, provided that the addict is *genuinely attracted* to taking the drug. And he tells us that a person experiences genuine attraction if that person “finds the occurrence of the event attractive or appealing, is enthusiastic about it (at least to some extent), and tends to view it with pleasure or gusto” (“Which Desires Are Relevant to Well-Being?” 674). Heathwood implicitly distinguishes *pleasure* from *gusto*, *enthusiasm*, and *finding things attractive or appealing*. I tend to think the distinction is somewhat misleading because all these states are often pleasant. Insofar as they are not—insofar as “gusto” is experienced as a pure “urge,” for example—I contend that they are not constitutive of welfare attitudes.

Vulcans' lives may be literally joyless, without the pursuit of pleasure or happiness to motivate them. They won't eat at fine restaurants to enjoy the food. But they may nevertheless have serious intellectual and moral goals. They may want to advance science, for example, and to help those around them. They might even want to build a family or make money. They experience no pleasure when anticipating or achieving these goals, but they value and pursue the goals all the same.⁴¹

Chalmers clearly thinks that Vulcans have welfare and illfare attitudes, in my sense. But hedonic subjectivists are committed to denying this.

The first thing to note is that we need to be careful in interpreting Chalmers's claim that Vulcans want things and value their goals. For as Chalmers himself writes, mental terms are often ambiguous.⁴² Vulcans can certainly want things and value their goals in a *functional* sense. Vulcans can designate certain states of affairs as their goals, and they can behave in complex and ingenious ways to bring it about that those state of affairs obtain. But that alone is not enough for a Vulcan to have welfare or illfare attitudes. After all, philosophical zombies can want things and value their goals in this purely functional sense, but few philosophers accept that zombies have welfare or illfare attitudes.⁴³

It is true that Vulcans, unlike zombies, have phenomenology. But this is not enough for them to have welfare or illfare attitudes. It is not enough to have *some phenomenology or other* in addition to having desires and values in the purely functional sense. For if we imagine beings that want and value things in a purely functional sense, while having phenomenology that is mismatched with their behavior, then it is clear that these beings do not have welfare or illfare attitudes either. To make the issue more concrete, imagine three subjects are collecting rare flowers. All of the subjects make sophisticated plans to collect the flowers, and all pursue those plans with roughly the same level of determination. They differ only with respect to their phenomenology: the first subject is an ordinary person, the second subject has mismatched phenomenology, and the third person is a Vulcan. As an ordinary person, the first subject tends to have the sorts of hedonic feelings that are characteristic of purposeful planning and action. For example:

1. They tend to feel mildly excited or hopeful at the thought of finding the flowers.

41 Chalmers, *Reality+*, 327.

42 Chalmers, *The Conscious Mind*, 11–22.

43 For discussions of zombie well-being or lack thereof, see Siewert, "Consciousness"; and Kriegel, "The Value of Consciousness to the One Who Has It."

2. They tend to feel mildly worried or gloomy at the thought of failing to find the flowers.
3. They tend to feel mild annoyance or frustration when their goals are thwarted.
4. They tend to feel mild excitement and enthusiasm when they advance their goals.
5. They tend to feel moderately gratified or proud when their plans come to fruition.

I am not imagining the search for the flowers is an emotional rollercoaster; I only suppose that this person cares about advancing their goals in an ordinary sort of way. Of course, to the extent that she succeeds in finding rare flowers, this is good for her.

The second of the subjects has phenomenology that is mismatched from their behavior. Perhaps they have only visual and auditory phenomenology: it is as if they are placidly watching a movie about collecting rare flowers, filmed from the perspective of the flower collector. They have no emotional investment in the events of the internal movie; they are merely a passive recipient of sensation. Since this subject behaves just like an ordinary person, they may smile when they succeed in finding a rare flower and frown when they fail, but this behavior is never indicative of their underlying feelings. They get excited about success to exactly the same degree that my toaster gets excited about toasting bread. They get worried about failure to exactly the same degree that my toaster gets worried about being unplugged. It appears as though the mismatched subject is genuinely invested in their project of finding rare flowers, but this is merely an appearance.

The third subject, the Vulcan, is supposed to occupy some sort of middle ground between the ordinary person and the mismatched subject. Like the mismatched subject, their planning and behavior does not involve any sort of hedonic ups and downs, but like the ordinary subject, their phenomenology is sufficient for their having welfare and illfare attitudes. So we have a kind of negative description of Vulcan phenomenology (no hedonic ups and downs) plus a description of the phenomenology in terms of its normative-theoretical role (it suffices for welfare and illfare attitudes). But this leaves the crucial question unanswered: What is the relevant phenomenology? Why should we think that any such phenomenology exists?⁴⁴

44 In a recent discussion, Luke Roelofs goes further in characterizing the nonhedonic but welfare-relevant phenomenology that Vulcans might have. They write that Vulcans may have *motivating consciousness*, which they describe as follows: "Motivating consciousness refers specifically to conscious states which participate on the conative side, by making

Chalmers might simply say that Vulcans have the phenomenology of desire, minus any feelings of pleasantness or unpleasantness. But I have already considered much of what might be thought to be the phenomenology of desire: the phenomenology of seeing something as good, the phenomenology of pure “wanting,” and various forms of mixed hedonic phenomenology. I argue that none of these kinds of phenomenology pose a problem for hedonic subjectivism. So if there is some other phenomenology that does not fall under any of these headings, then the onus is on Chalmers and other friends of Vulcans to identify it. Until then, we ought to doubt that Vulcans have welfare and illfare attitudes. We should think that they are like a mismatched subject. They have phenomenology, they may plan and act as though they genuinely care about things, but they do not genuinely care.

5. THE IMPORTANCE OF PLEASURE AND DISPLEASURE

The hedonic view accounts for the distinction between welfare and illfare attitudes in a highly straightforward way. It avoids important problems for hedonism and desire-based versions of subjectivism, and it is not defeated by the counterexamples I have considered here. I conclude that it ought to be taken seriously as an account of the subjective component of well-being.⁴⁵

Lingnan University
palliesdan@gmail.com

REFERENCES

- Ayars, Alisabeth. “Attraction, Aversion, and Meaning in Life.” *Journal of Ethics and Social Philosophy* 28, no. 3 (2024): 386–410.

some option, outcome, or action appear good or attractive” (“Sentientism, Motivation, and Philosophical Vulcans,” 316). The question for Roelofs is essentially the same as the question for Chalmers: *What is this phenomenology?* Once we have a positive conception of it, we can then evaluate the proposal that having this phenomenology is sufficient for having welfare attitudes.

- 45 I worked on this paper, in various forms, for a long time, and I have received helpful feedback from a great many people. Many thanks to Janet Levin, Ralph Wedgwood, John Hawthorne, Chris Heathwood, Uriah Kriegel, Jennifer Foster, Alex Dietz, James Fanciullo, Adam Bradley, Jesse Hill, as well as audiences at the University of Southern California Speculative Society, the Value of Consciousness series, and two anonymous reviewers at *JESP*. Special thanks to Mark Schroeder for his encouragement, support, and many rounds of detailed comments.


- Barlassina, Luca, and Max Khan Hayward. "More of Me! Less of Me! Reflexive Imperativism About Affective Phenomenal Character." *Mind* 128, no. 512 (2019): 1013–44.
- Berridge, Kent C. "Wanting and Liking: Observations from the Neuroscience and Psychology Laboratory." *Inquiry* 52, no. 4 (2009): 378–98.
- Berridge, Kent, and Piotr Winkielman. "What Is an Unconscious Emotion? The Case for Unconscious 'Liking'." *Cognition and Emotion* 17, no. 2 (2003): 181–211.
- Chalmers, David John. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press, 1997.
- . *Reality+: Virtual Worlds and the Problems of Philosophy*. W. W. Norton, 2022.
- Chang, Ruth. "Can Desires Provide Reasons for Action?" In *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, edited by R. Jay Wallace, Philip Pettit, Samuel Scheffler, and Michael Smith. Clarendon Press, 2004.
- Dorsey, Dale. "Subjectivism Without Desire." *Philosophical Review* 121, no. 3 (2012): 407–42.
- Fanciullo, James. "Alienation, Engagement, and Welfare." *Philosophical Quarterly* 75, no. 1 (2025): 40–60.
- Feldman, Fred. *Pleasure and the Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism*. Clarendon Press, 2004.
- . "Two Questions About Pleasure." In *Philosophical Analysis*, edited by D. F. Austin. Kluwer Academic Publishers, 1988.
- Fletcher, Guy. "A Fresh Start for the Objective-List Theory of Well-Being." *Utilitas* 25, no. 2 (2013): 206–20.
- Gosling, J. C. B. *Pleasure and Desire: The Case of Hedonism Reviewed*. Oxford University Press, 1969.
- Gregory, Alex. *Desire as Belief: A Study of Desire, Motivation, and Rationality*. Oxford University Press, 2021.
- Heathwood, Chris. "Desire-Fulfillment Theory." In *The Routledge Handbook of Philosophy of Well-Being*, edited by Guy Fletcher. Routledge, 2015.
- . "Ill-Being for Desire Satisfactionists." *Midwest Studies in Philosophy* 46 (2022): 33–54.
- . "Which Desires Are Relevant to Well-Being?" *Noûs* 53, no. 3 (2019): 664–88.
- Kagan, Shelly. "An Introduction to Ill-Being." *Oxford Studies in Normative Ethics* 4 (2014): 261–88.
- Kelley, Anthony Bernard. "Well-Being and Alienation." PhD diss., University of Colorado, 2020. ProQuest (2476324285).
- Kriegel, Uriah. *Subjective Consciousness: A Self-Representational Theory*. Oxford

- University Press, 2009.
- . “The Value of Consciousness to the One Who Has It.” In *The Importance of Being Conscious*, edited by Geoffrey Lee and Adam Pautz. Oxford University Press, forthcoming.
- Kringelbach, Morten L., and Kent C. Berridge. “The Joyful Mind.” *Scientific American* 307, no. 2 (2012): 40–45.
- Lewis, David. “Psychophysical and Theoretical Identifications.” *Australasian Journal of Philosophy* 50, no. 3 (1972): 249–58.
- Lin, Eden. “How to Use the Experience Machine.” *Utilitas* 28, no. 3 (2016): 314–32.
- . “Two Kinds of Desire Theory of Well-Being.” *Midwest Studies in Philosophy* 46 (2022): 55–86.
- Lovett, Adam, and Stefan Riedener. “The Good Life as the Life in Touch with the Good.” *Philosophical Studies* 181, no. 5 (2024): 1141–65.
- Mathison, Eric. “Asymmetries and Ill-Being.” PhD diss., University of Toronto, 2018.
- Millikan, Ruth Garrett. *Language: A Biological Model*. Clarendon Press, 2005.
- Murphy, Mark C. *Natural Law and Practical Rationality*. Cambridge University Press, 2001.
- Nozick, Robert. *Anarchy, State, and Utopia*. Basic Books, 1974.
- Oddie, Graham. *Value, Reality, and Desire*. Clarendon Press, 2005.
- Pallies, Daniel. “Attraction, Aversion, and Asymmetrical Desires.” *Ethics* 132, no. 3 (2022): 598–620.
- . “An Honest Look at Hybrid Theories of Pleasure.” *Philosophical Studies* 178, no. 3 (2021): 887–907.
- Papineau, David. “Representation and Explanation.” *Philosophy of Science* 51, no. 4 (1984): 550–72.
- Parfit, Derek. *Reasons and Persons*. Oxford University Press, 1986.
- Peciña, Susana, Kyle S. Smith, and Kent C. Berridge. “Hedonic Hot Spots in the Brain.” *Neuroscientist* 12, no. 6 (2006): 500–11.
- Quinn, Warren. “Putting Rationality in Its Place.” In *Value, Welfare, and Morality*, edited by Christopher W. Morris and R. G. Frey. Cambridge University Press, 1993.
- Raibley, Jason R. “Values, Agency, and Welfare.” *Philosophical Topics* 41, no. 1 (2013): 187–214.
- Rawls, John. *A Theory of Justice*. Belknap Press, 1971.
- Roelofs, Luke. “Sentientism, Motivation, and Philosophical Vulcans.” *Pacific Philosophical Quarterly* 104, no. 2 (2023): 301–23.
- Scanlon, Thomas. *What We Owe to Each Other*. Belknap Press, 1998.
- Schroeder, Mark. *Slaves of the Passions*. Oxford University Press, 2007.

- Schroeder, Timothy. *Three Faces of Desire*. Oxford University Press, 2004.
- Siewert, Charles. "Consciousness: Value, Concern, Respect." In *Oxford Studies in Philosophy of Mind*, vol. 1, edited by Uriah Kriegel. Oxford University Press, 2021.
- Sinhababu, Neil. *Humean Nature: How Desire Explains Action, Thought, and Feeling*. Oxford University Press, 2017.
- Smithies, Declan. "A Hedonic Theory of Desire." *Australasian Journal of Philosophy* (forthcoming).
- Smithies, Declan, and Jeremy Weiss. "Affective Experience, Desire, and Reasons for Action." *Analytic Philosophy* 60, no. 1 (2019): 27–54.
- Stampe, Dennis W. "The Authority of Desire." *Philosophical Review* 96, no. 3 (1987): 335–81.
- Strawson, Galen. "Self-Intimation." *Phenomenology and the Cognitive Sciences* 14, no. 1 (2013): 1–31.
- Tiberius, Valerie. *Well-Being as Value Fulfillment: How We Can Help Each Other to Live Well*. Oxford University Press, 2018.
- Wall, Steven, and David Sobel. "A Robust Hybrid Theory of Well-Being." *Philosophical Studies* 178, no. 9 (2020): 2829–51.
- Williams, Bernard. "Internal and External Reasons." In *Rational Action*, edited by Ross Harrison. Cambridge University Press, 1979.

FREEDOM OF GENDER

Rach Cosker-Rowland

N MAY 29, 2020, a change to the Hungarian Registry Act came into place that made it impossible for trans people to legally change their gender.¹ Because of this law, trans people in Hungary now cannot change their legal gender from the gender that they were assigned at birth to the gender that matches their gender identity and cannot change their gender markers on their official legal documents, such as their passports, to match their gender identity. Furthermore, Hungary has two distinct official lists for the names of men and women. Trans people now cannot adopt and use a name on their legal documents that is on the list for the gender they were not assigned at birth.² So trans women in Hungary cannot change their name so that a woman's name appears on their identity documents (such as their driving licenses) and cannot change their gender marker on their legal documents (such as their passports): trans women who did not change their legal gender prior to May 29, 2020, will forever have Hungarian passports and legal names that label them as men. Some US states have recently followed in Hungary's footsteps. In 2023 and 2024, Kansas, Montana, North Dakota, Tennessee, Oklahoma, Florida, and Texas all made it impossible for trans people in these states to change their gender markers on their birth certificates and driving licenses.³ And in early 2025, the US federal government made it impossible for trans people to renew or get new passports

1 See Faye, *The Transgender Issue*, 160.

2 See Andersson, "I Won't Even Be Allowed to Use My Name Now that Hungary Has Scrapped All Rights for Trans People."

3 See Reed, "Anti-Trans Legislative Risk Assessment Map," "Kansas, Other States Threaten to Undo Legal Gender Changes," and "Tennessee Law Rolls Back Trans Rights, Regressively Defines Sex"; Hanna, "Kansas Attorney General Sues to Prevent Transgender People from Changing Driver's Licenses"; Lang, "Texas Just Quietly Revoked Trans People's Ability to Change Their Birth Certificates"; Rummler, "Florida Is Quietly Denying Transgender Residents Updated Birth Certificates"; Yurcaba, "Florida Bars Transgender People from Changing the Sex on Their Driver's Licenses"; and "Identity Document Laws and Policies" by the Movement Advancement Project, https://www.lgbtmap.org/equality-maps/identity_documents/birth_certificate (regularly updated).

that label them as the gender that matches their gender identity rather than the gender they were assigned at birth.⁴

Trans people who are unable to change their name or gender marker on their identification documents face many harms. We have to present our identification documents all the time: to collect mail and prescriptions, at airports, when setting up a bank account or renting a house, when buying alcohol, entering bars or music venues, or when requested by the police. Many trans people appear to be the gender that matches their gender identity rather than the gender they were assigned at birth: many trans women are socially perceived to be women rather than men; many trans men are socially perceived to be men. A trans person showing an identity document that presents them as the gender they were assigned at birth rather than the gender they take themselves to be, present as, and are taken to be a member of by others often outs them as trans. And trans people are widely stigmatized and subject to abuse and attack.⁵ So because trans people's showing an identity document on which their gender does not match their gender identity and presentation often outs them as trans, trans people's having to use such identification documents subjects them to abuse. According to the 2015 US Transgender Survey, 25 percent of trans people have experienced verbal harassment because they have had to use identification documents that present them as the gender they were assigned at birth.⁶

This article asks whether trans people have moral rights to change their gender markers on their legal identification documents. It argues that trans people have rights to relatively easily ensure that their gender markers on their legal identification documents do not clash with their gender identities and that these rights have a particularly stringent grounding—namely, they are grounded in basic liberal rights to live and act with integrity.

It is useful to clarify some issues about sex and gender as they are understood in this article and by legal systems. I assume that we can distinguish between sex and gender. However, in many legal systems such as the United Kingdom and Australia, there is only a category of legal sex/legal gender, and sex and gender are conflated and grouped into a single category by the law. For instance, in a document explaining that its passports present gender information rather

4 See Hansler, "State Department Suspends Processing Passport Applications with 'X' Marker"; Theil, "Trans Americans Accuse Trump of 'Travel Ban'"; and CNN, "This Trans Influencer Received a Passport with the Wrong Gender After Trump's Executive Order."

5 See James et al., "The Report of the 2015 US Transgender Survey"; Faye, *The Transgender Issue*, esp. 166–67; Trevor Project, "National Survey of LGBTQ Youth Mental Health (2020)"; Reed, "Tennessee Law Rolls Back Trans Rights, Regressively Defines Sex"; and Weiss, "The Gender Caste System," 133, 150, 152, 168.

6 James et al., "The Report of the 2015 US Transgender Survey," 89.

than sex information despite using the categories ‘male’, ‘female’, and ‘X’, the Australian government notes that “although sex and gender are conceptually distinct, these terms are commonly used interchangeably, including in legislation.”⁷ In the United Kingdom, one is often taken to have only a legal sex, not a legal gender, but as Davina Cooper and colleagues explain, ‘legal sex’ and ‘legal gender’ are often used interchangeably in UK contexts, and this is true in government documents.⁸ For instance, the UK government states that one’s “affirmed gender” will be legally recognized (e.g., a trans man will be legally recognized as a man) if one has a gender recognition certificate; and a gender recognition certificate is the certificate one needs to change the marker on one’s birth certificate from male to female or vice versa.⁹ It might seem that we are not assigned a gender at birth but are only assigned a sex on our birth certificates. However, some jurisdictions such as Australia explicitly reject this claim.¹⁰ And many other jurisdictions (including the United Kingdom and most US states) allow trans men, for instance, to change their birth certificates and passports to present them as men rather than women because they have male gender identities and live as men—not because they have changed from being biologically female to being biologically male.¹¹

This article does not rely on claims about the metaphysics of sex or gender. It does rely on the idea that an adult being presented as female on their identification documents presents them as a woman. There are several reasons for this. For instance, first, as discussed above, many jurisdictions take their female and male gender markers on documents to denote gender, not sex. Second, ‘female’ and ‘male’ are often used as synonyms for ‘woman’ and ‘man’, and in part because of this (and in part because of other views they have), many, and perhaps most, people will assume that an adult is a woman if they are presented as female on their documents or that they are a man if they are presented as male on these documents.¹² In general, this article aims to stay neutral on what

7 See Australian Government, “Guidelines on the Recognition of Sex and Gender.”

8 Cooper et al., “Abolishing Legal Sex Status,” 91.

9 See the UK government webpage “Apply for a Gender Recognition Certificate,” <https://www.gov.uk/apply-gender-recognition-certificate/> (updated November 26, 2024; accessed August 21, 2025).

10 Australian Government, “Guidelines on the Recognition of Sex and Gender,” 4.

11 See, e.g., the UK government webpage “Apply for a Gender Recognition Certificate” (above n. 9); and, for discussion, Hines, *Gender Diversity, Recognition and Citizenship*.

12 This synonymous use is present even in (trans-inclusive) philosophical literature on gender, where, for instance, ‘female gender identity’ generally appears as a synonym for ‘the gender identity “woman”’ because it is sometimes easier to use. See, e.g., Jenkins, “Amelioration and Inclusion,” 404, 410.

gender is, for this metaphysical issue is irrelevant to the question of what our moral rights are regarding the gender on our legal documents.¹³

Section 1 sketches our basic liberal rights to live and act with integrity, arguing that given that we have such basic liberal rights to live and act with integrity, we have *pro tanto* rights to *freedom of legal gender identification*—that is, *pro tanto* rights to change our legal documents so that they do not clash with our gender identities.

Section 2 argues that these *pro tanto* rights to freedom of legal gender identification give rise to all-things-considered rights to freedom of legal gender identification; these *pro tanto* rights are not outweighed by other considerations. In this case, as I explain, blanket bans on gender marker changes, such as Hungary's, Kansas's, North Dakota's, Montana's, Oklahoma's, Florida's, Texas's, Tennessee's, and the US government's, are unjust and breach trans people's all-things-considered rights. Section 2 also discusses policies that require that trans people have sex reassignment surgery before they can change their gender markers and policies that do not allow nonbinary people to have X markers (rather than markers denoting male or female) on their passports. I argue that these policies breach trans and nonbinary people's all-things-considered rights too. Section 2 then argues that, in light of these arguments, there is a strong presumption in favor of one of several policies that make it relatively easy for trans and nonbinary people to have legal identification documents that do not clash with their gender identities.

The topic of trans people's rights to change their gender markers on their legal identification documents has received little attention in philosophy. And the few philosophical discussions of this topic have been critical of such rights: philosophers Holly Lawford-Smith and Kathleen Stock both argue that trans people do not have rights to change their genders on their legal identification documents because of the consequences of granting trans people these rights. Section 2 discusses these arguments and shows that they fail.

Beyond philosophy, the topics of gender marker change on legal documents and of the removal of gender markers from these documents altogether have been discussed by activists, human rights groups, academic lawyers, and legal theorists. Much of this discussion is primarily concerned with (1) the legal rights that trans people have in particular jurisdictions (and their legal grounds) or with (2) the virtues and vices of particular law reforms, rather than with (3) the moral rights that trans people have or how we should think about the moral

13 For similar claims and arguments, see Jenkins, *Ontology and Oppression*, ch. 8; and Dembroff, "Real Talk on the Metaphysics of Gender."

grounds of these rights.¹⁴ For instance, the Yogyakarta Principles outline a set of principles for the application of international human rights law to sexual orientation and gender identity, and one of those principles requires states to ensure that “all state-issued identity papers which indicate a person’s gender/sex—including birth certificates, passports, electoral records and other documents—reflect the person’s profound self-defined gender identity.”¹⁵ But the Yogyakarta Principles do not discuss the grounds of trans people’s rights to have their identification documents reflect their gender identities.¹⁶ Some of this literature beyond philosophy is also somewhat skeptical that it is advantageous to think of these issues in a rights-based framework.¹⁷ However, we can glean several alternative candidates for the grounds of rights to freedom of legal gender identification from this literature. Section 3 discusses such alternative (harm-based, privacy-based, and autonomy-based) grounds of our rights to freedom of legal gender identification and argues that there are virtues to adopting an integrity-based approach over these alternatives.

1. INTEGRITY-BASED RIGHTS TO FREEDOM OF LEGAL GENDER IDENTIFICATION

1.1. Basic Liberal Rights to Live and Act with Integrity

Basic liberal rights include rights to freedom of religious belief and expression and freedom of political speech.¹⁸ Basic liberal rights are not unlimited. We have a basic liberal right to freedom of speech, but this does not mean that

14 On 1, see, e.g., Cannoot and Decoster, “The Abolition of Sex/Gender Registration in the Age of Gender Self-Determination”; Hines, *Gender Diversity, Recognition and Citizenship*; and Weiss, “The Gender Caste System.” On 2 see, e.g., Ashley, “‘X’ Why?”; Cooper and Emerton, “Pulling the Thread of Decertification,” 5–9; and Neuman Wipfler, “Identity Crisis.” However, for examples of brief discussion of the moral grounds of the right to have legal documents match one’s gender identity, see Lau, “Gender Recognition as a Human Right,” esp. 194–95; and Pearce et al., “Introduction,” 15. See further section 3 below.

15 The Yogyakarta Principles: Principles on the Application of International Human Rights Law in Relation to Sexual Orientation and Gender Identity (March 2007), https://yogyakartaprinciples.org/wp-content/uploads/2016/08/principles_en.pdf, 12.

16 Its principles regarding nondiscrimination and privacy are not connected to the right to not be forced to have identification documents that are out of line with one’s self-defined gender. See the Yogyakarta Principles, 10, 14.

17 See, e.g., Renz, “Genders that Don’t Matter”; Venditti, “Gender Kaleidoscope,” 72; and Spade, *Normal Lie*, ch. 3, 93. My argument in this article shows that we should not be skeptical in this way, since basic liberal rights establish that there are such rights to freedom of legal gender identification.

18 See, for instance, Dworkin, *A Matter of Principle*, esp. 191–92, and *Justice for Hedgehogs*, 371.

we have a right to yell "Fire!" in a crowded theater or a right to whip up a mob in front of someone's house, inciting them to burn it down; and according to many, our rights to freedom of speech do not establish that we have a right to spout hate speech. These basic liberal rights are codified in human rights conventions such as the United Nations Universal Declaration of Human Rights (1948) and the European Convention on Human Rights (1950). And some form of recognition and protection of these basic liberal rights is often taken to be a necessary condition on—or even constitutive of—a regime or political philosophical view being liberal.¹⁹

A state that forces all of its citizens to go to mass breaches basic liberal rights to freedom of religious belief and expression. In Malaysia, ethnic Malays are assigned Muslim at birth: they have 'Muslim' presented on their legal identification documents and are expected to engage in Islamic fasting and prayer—and can be fined for refraining from doing so. Ethnic Malays are assigned Muslim regardless of whether their families are Muslim or whether they ever practice Islam or ever believe the tenets of any form of Islam. In most Malaysian states, it is impossible to change one's legal status from Muslim to another religion or to no religion. In other Malaysian states, changing one's legal religion from Muslim is in principle possible, though only after months or years in a reeducation center; and in practice, most of those assigned Muslim at birth who wish to change their legal religion struggle to do so.²⁰ Malaysia breaches ethnic Malays' basic liberal rights to freedom of religious belief and expression. Other famous examples of breaches of freedom of religious belief and expression include polytheists being forced to pledge allegiance to a monotheistic God, Jehovah's Witnesses being forced to pledge allegiance to the US flag in US schools even though they take doing this to involve a wrongful form of idolatry, and Seventh Day Adventists being forced to work on their holy day (or else face poverty without pay or unemployment support).²¹ I presume that it is unjust for a state to breach its citizens' freedom of religious belief and expression in these ways. And so I presume that basic liberal rights to freedom of religious belief and expression at least protect our religious belief and expression in these ways.

Many political liberals take these basic liberal rights to freedom of religious belief and expression to be grounded in more general rights to live a life that we take to be right, meaningful, or good. According to John Rawls, we have two

19 See, for instance, Dworkin, *Justice for Hedgehogs*, 371, and *Religion Without God*, 105–6.

20 See Ahmad et al., "Freedom of Religion and Apostasy"; Chen, "Renouncing Islam in Malaysia Is Dangerous"; Aziz, "Freedom of Religion by Religion for Religion"; Samuri and Quraishi, "Negotiating Apostasy"; and Nazri, "What Happens when Muslims in Malaysia Try to Leave Islam" (including the references therein).

21 See Nussbaum, *Liberty of Conscience*, 135–36, 204–14.

moral powers: a rational power to form, revise, and pursue a conception of the good or valuable life for us; and a reasonable power to form, revise, and pursue a conception of what is right and wrong.²² Rawls holds that we need liberal rights to freedom of religious belief and expression in order to exercise these powers: in order to be able to form and revise conceptions of the good life for us and what is right, we need to freely discuss different religious and nonreligious views of the good life and of what is right; and we need to be able to act in line with our conceptions of the right and the good in order to exercise our powers to pursue our conceptions of the right and the good.²³ Ronald Dworkin, Jocelyn Maclure, Charles Taylor, and Robert Audi hold similar views, according to which our rights to freedom of religious belief and expression are grounded in more general rights to live a life that we take to be right, meaningful, or good.²⁴ On this view, basic liberal rights protect atheist conscientious objectors' rights not to be forced to go to war just as much as they protect religious minorities' rights not to have to pledge allegiance to a God they do not believe in.

This Rawlsian view of the grounds of basic liberal rights to freedom of religious belief and expression is given its most comprehensive articulation in Cecile Laborde's *Liberalism's Religion*. Laborde holds that the best way of understanding this Rawlsian view involves understanding basic liberal rights to freedom of religious belief and expression to be grounded in general rights to live and act with integrity. Laborde, following Bernard Williams, holds that someone acts and lives with integrity when they act in line with

1. their practical identities;
2. their views of what gives their life meaning or what their life is "fundamentally about";
3. their views, commitments, or ideals regarding the kind of person they should be; and
4. the way of life that they value, and take to be good for them, though not necessarily for everyone else.²⁵

Integrity is a coherence notion on this picture. We have views or make judgments about what our acting in line with our practical identities, our ideal of how we should be or act, or our views of the good or meaningful life for us involves. And we act with integrity to the extent that we act in line with these

22 Rawls, *Justice as Fairness*, 45.

23 Rawls, *Justice as Fairness*, 45.

24 See Dworkin, *Justice for Hedgehogs*, 368–70; Maclure and Taylor, *Secularism and Freedom of Conscience*, 75–81; and Audi, "Religious Liberty Conceived as a Human Right," 418.

25 Laborde, *Liberalism's Religion*, 204–5. See also Williams, "A Critique of Utilitarianism," 108–18.

judgments and views. These ethical and moral views are not mere preferences. According to Laborde, unlike preferences, someone cannot act in a way that is out of line with these ethical and moral views without feeling negative reactive attitudes such as shame, remorse, or guilt.²⁶ So, according to Laborde:

Integrity-Based Rights: We have fundamental liberal rights to be able to live and act with integrity—that is, to live in line with our view of the life we ought to live, to live in line with our view of what the good or meaningful life is for us, or to live in line with our practical identities.²⁷

This view generates the right results in the cases that we started off with. For instance, many judge that the good life for them does not involve pledging allegiance to a God that they do not believe in or attending mass, and many judge that they ought not work on their holy day. So the importance to us of living with integrity—which is something that everyone can agree is important—grounds these rights to religious belief and expression.²⁸

These integrity-based rights are, at least normally, only negative rights. Suppose your ideal of the good and meaningful life for you involves going on a pilgrimage to a faraway important religious site; yet your fundamental liberal rights to live with integrity do not entitle you to all the resources you need to go on this pilgrimage.

Our rights to live and act with integrity are strong and important rights. Nonetheless, they are *pro tanto* rather than all-things-considered rights: if acting with integrity involves you encroaching on someone else's basic liberal rights, you do not have an all-things-considered right to act in that way. So by granting that we have integrity-based rights that ground our rights to freedom of religious belief and expression, we need not grant that all things that someone may feel entitled to as part of their freedom of religion are really things that they have a right to. For instance, many religious schools claim that they have a right to exclude LGBTQ+ children. We can hold that there are integrity-based rights but challenge the claim that religious schools have such a right in various ways: we can ask whether it is really true that in order to live with integrity anyone needs to not teach or receive schooling with LGBTQ+ children.²⁹ And if someone genuinely needs to not engage with LGBTQ+ people in order to live

26 Laborde, *Liberalism's Religion*, 204.

27 See Laborde, *Liberalism's Religion*. For similar views, see Bou-Habib, "A Theory of Religious Accommodation"; Dworkin, *Sovereign Virtue*, 270; and Billingham, "How Should Claims for Religious Exemptions Be Weighed?"

28 On how everyone can agree that this is important, see Laborde, *Liberalism's Religion*, 61; and Bou-Habib, "A Theory of Religious Accommodation," esp. 120.

29 See Sunstein, "On the Tension Between Sex Equality and Religious Freedom," 136–37.

with integrity, we might argue that they do not have an all-things-considered right to do this because of the message our state would send by allowing them to segregate themselves from LGBTQ+ people and the harm that this message would cause to LGBTQ+ people.³⁰

I discuss alternatives to the integrity-based approach to rights to freedom of religious belief and expression in subsection 1.3 below. But in the next subsection, I first explain how if we have rights to live and act with integrity, these rights generate *pro tanto* rights to gender marker change.

1.2. Freedom of Legal Gender Identification

We can distinguish between the following.

Gender Identity (GID): This is, most generally, the gender that you take yourself to be, your sense of what gender you are, or the gender category that seems to you to fit you.³¹

Assumed Gender (AG): This is the gender that we are assumed to be by strangers, such as those to whom we present identification documents (that is, whether we are taken by others to be a woman, a man, nonbinary or some particular nonbinary gender).

Documented Gender (DG): This is the gender that is listed and presented on our identification documents.³²

Suppose that Alexa is a trans woman whose GID and AG are “woman” and whose DG is “man.” There are what we might think of as intrinsic and extrinsic effects of the conflict between (1) Alexa’s DG and (2) her GID and AG.

Extrinsic Effects: This conflict can lead to people treating Alexa as a gender that clashes with Alexa’s GID (e.g., treating her as a man).

Intrinsic Effects: By presenting her gender documents, Alexa presents herself as a gender that clashes with her gender identity (e.g., as a man).

30 For discussion of how such messages can limit permissible actions and policies, see, e.g., Lever, “Why Racial Profiling Is Hard to Justify,” 97; and Hellman, “Racial Profiling and the Meaning of Racial Categories,” 237.

31 See Stryker, *Transgender History*, 21; Jenkins, “Amelioration and Inclusion,” 409; Bettcher, “Through the Looking Glass,” 396; and Cosker-Rowland, “Gender Identity.”

32 I do not take our GID, AG, and DG to be all that matters to us, or even what matters most to us about sex and gender, and certainly not to be all there is to sex and gender. I make and use these distinctions for the purpose of explaining how restrictions on gender marker change impinge on the integrity of trans people.

These intrinsic and extrinsic effects of a clash between trans people's (1) DG and (2) GID and AG can impinge on their integrity in several ways.

1.2.1. *Practical Identity*

To be trans is to have a gender identity that is different from the gender you were assigned at birth. For instance, trans women were assigned male at birth but have female gender identities. (They have the gender identity 'woman'.)³³ To be trans and have the gender identity 'woman' involves seeing it as an important part of your identity that you live as, or should live as, a woman.³⁴ In order to live with integrity, we need to live and act in line with our practical identities. The intrinsic and extrinsic effects of the conflict between Alexa's documented gender and her female gender identity impinge on her ability to live as a woman: she has to present herself as a man rather than a woman whenever she produces her legal identification documents, and doing this will lead to her being treated by others as a man rather than as a woman. So Alexa's inability to change her documented gender to 'woman' impinges on Alexa's integrity by impinging upon her ability to live and act in line with one of her practical identities.

1.2.2. *Ought Judgments*

On many accounts of gender identity, trans gender identities involve normative judgments. According to Katharine Jenkins's account of gender identity, to have a trans gender identity is to judge or to have the experience that you ought to navigate the world, categorize yourself, and be categorized in line with norms associated with a gender other than the gender that you were assigned at birth.³⁵ According to Susan Stryker, gender identities involve our judgments about the gender category that fits us or that it is correct for us to present ourselves as or be treated as.³⁶ Many trans people discuss their first experiences of their gender identity and gender in terms of their normative experiences. For instance, trans women Julia Serano and Mia Violet explain how it seemed to them that they *ought* to line up with the girls rather than the boys when they were at school and *should* use the girls' bathrooms rather than the boys' bathrooms.³⁷ Many trans people also discuss and understand their genders in terms of the gender

33 See Stryker, *Transgender History*, ch. 1; and Faye, *The Transgender Issue*, xiv.

34 See Bettcher, "Evil Deceivers and Make-Believers," 46; and Barnes, "Gender and Gender Terms," 709.

35 Jenkins, "Amelioration and Inclusion," 411, 413.

36 Stryker, *Transgender History*, 21. See also Cosker-Rowland, "Gender Identity."

37 Serano, *Whipping Girl*, 78; and Violet, *Yes, You Are Trans Enough*, 24.

categories that fit them or that it is correct to categorize them as.³⁸ And if it is fitting or correct to ϕ , other things equal, one ought to ϕ . For example, if it is fitting or correct for me to admire someone, then other things equal, I ought to admire them.³⁹ According to the account of integrity explained in section 1.1, if I judge that I ought to live as gender G or that my ideal of the life I ought to live involves living as a G , then in order to live with integrity, I must live as a G . Being unable to change our documented gender so that it does not clash with our gender identity and with the gender that we judge we ought to live as impinges on our ability to live as a gender that does not clash with our gender identity or the gender that we judge we ought to live as. For if we are unable to change our documented gender so that it does not clash with our gender identity, we must present ourselves as a gender that clashes with our gender identity to access many goods and services, and this is likely to lead to others frequently treating us as a gender that clashes with our gender identity. So a trans person's being unable to change their documented gender so that it does not clash with their gender identity and the gender that they judge they ought to live as impinges on their ability to live and act with integrity.

1.2.3. Authenticity

Many trans people judge that in order to be authentic, they need to live their lives as a gender that is different from the gender they were assigned at birth; they take it to be inauthentic to live as the gender they were assigned at birth.⁴⁰ Plausibly, to be trans involves judging that you are a gender other than that which you were assigned at birth and that your being authentic involves living as that gender.⁴¹ If you judge that in order to be authentic, you need to live your life as gender G , to treat yourself and be treated as a G , then you take your living your life as who you really are to involve living as a G , or you take it that what your life is fundamentally about or your living a life that is meaningful for you involves living as gender G . And if you take living your life as who you really are to involve living as a G , or you take it that what your life is fundamentally about or your living a life that is meaningful for you to involve your living as gender G ,

38 See Roche, *Trans Power*, 16–17; Weiss, “9 Things People Get Wrong About Being Nonbinary”; and Rajunov and Duane, *Nonbinary*, 28, 94, 109, 231.

39 See Cosker-Rowland and Howard, “Fittingness,” 4–5.

40 See Kee, “35 People Who Transitioned on How It Impacted Their Mental Health”; Williams, “What It Means to Be Authentic”; Cook, “10 Transgender People Share What They Wish They Knew Before Transitioning”; Violet, “The Fact I Can’t Marry as a Bride Is Another Reminder of How Unequal Trans Rights Still Are”; and Cosker-Rowland, “Integrity and Rights to Gender-Affirming Healthcare,” 833–34.

41 See Stryker, *Transgender History*, 21; and Bettcher, “Through the Looking Glass,” 396.

then you need to live as a G in order to live with integrity.⁴² (See section 1.1.) A trans person's being unable to change their documented gender so that it does not clash with the gender they take it that their authenticity requires them to live as impinges on their ability to live as that gender. So a trans person's being unable to change their documented gender so that it does not clash with the gender they take it that their authenticity requires them to live as impinges on their ability to live and act with integrity.

1.2.4. *Reactive Attitudes*

Many trans and nonbinary people experience guilt, shame, or other negative reactive attitudes towards themselves or judge that they are living wrongly by living as the gender they were assigned at birth.⁴³ For instance, many trans men feel ashamed that they have not transitioned to live as men because they judge that they are not being true to themselves or living as who they really are because they are not living as men; they are hiding who they are by refraining from transitioning. As discussed in section 1.1, if one has such negative reactive attitudes towards one's not ϕ -ing, then not ϕ -ing clashes with one's integrity. Trans people's being unable to change their documented gender so that it is different from the gender they were assigned at birth impinges on their ability to live as a gender different from that which they were assigned at birth. For trans people who are unable to change their documented gender will have to present themselves as a gender that clashes with their gender identity to access many goods and services—and so, to that extent, will have to live as a gender that clashes with their gender identity. And this will likely lead to others frequently treating them as a gender that clashes with their gender identity—and so, to an extent, will lead to their not living in line with their gender identity. So trans people's being unable to change their documented gender so that it is different from the gender they were assigned at birth impinges on their ability to live with integrity.

1.2.5. *What a Good and Meaningful Life for One Involves*

Many trans people judge that in order to live a good or meaningful life, or a life in which they can be happy, they must live as and be treated as the gender that matches their gender identity; and many trans people are very unhappy and find it impossible to live a good life while living as the gender they were

42 I argue elsewhere ("Integrity and Rights to Gender-Affirming Healthcare," 833–34) that if you need X in order to live authentically, then you need X in order to live with integrity.

43 See, e.g., Kee, "35 People Who Transitioned on How It Impacted Their Mental Health"; Rachel's story in Brighter, "Trans-Later"; Giordano, "Understanding the Emotion of Shame in Transgender Individuals"; and Serano, *Whipping Girl*, 78.

assigned at birth.⁴⁴ If one judges that one's living a good, valuable, or meaningful life involves one's living as gender *G*, then one's living with integrity involves one's living as a *G*. Trans people's being unable to change their documented gender so that it does not clash with their gender identity impinges on their ability to live as the gender that matches their gender identity. And so trans people's being unable to change their documented gender so that it does not clash with their gender identity impinges on their ability to live with integrity because it impinges on their ability to live a life that they take to be a good, valuable, or meaningful one for them.

1.2.6. *Misrepresentation*

There is a final slightly more indirect way in which trans people's being unable to change their documented gender so that it does not clash with their gender identity impinges on their ability to live with integrity. Many trans people judge that they are misrepresenting themselves or presenting themselves as someone who they are not by presenting themselves as the gender they were assigned at birth rather than the gender they take themselves to be.⁴⁵ And many people would judge that a good or valuable life for them, or a life in which they are living in line with their ideals of how they ought to live or are living authentically, is not a life in which they are frequently misrepresenting themselves or who they are to others. So, many people need to not be frequently misrepresenting themselves to others in order to live with integrity. But if one cannot change one's documented gender so that it does not clash with the gender one believes oneself to be, one cannot avoid frequently having to misrepresent oneself to others. So, for many people, their being unable to change their documented gender so that it does not clash with the gender they take themselves to be would stop them from living with integrity, or at least impinge on their ability to live with integrity. This is the ground on which intersex and nonbinary people were granted the right to have *X* gender markers on their US passports. In 2021, Dana Zzyym became the first US citizen to have an *X* gender marker (rather than an *M* or *F* marker) on their passport. The judge in Zzyym's case argued that nonbinary and intersex people ought to be able to have *X* markers on their passports because they have the right to travel internationally without lying to others about their gender, without presenting themselves as a gender that they take to be a lie.⁴⁶

44 See, e.g., Oladipo, "Majority of Trans Adults Are Happier After Transitioning"; and Violet, "The Fact I Can't Marry as a Bride Is Another Reminder of How Unequal Trans Rights Still Are."

45 See, e.g., Renz, "Genders that Don't Matter," 10.

46 See Clarke, "They, Them, and Theirs," 919.

If the arguments that I have made in this section are sound, then our basic liberal rights to live and act with integrity establish (*pro tanto*) rights to freedom of legal gender identification—that is, *pro tanto* rights for trans people to be able to change their documented gender so that it does not clash with their gender identity.

1.3. *Objections*

One objection to my argument that there are *pro tanto* rights to freedom of legal gender identification is that we do not have basic liberal rights to live and act with integrity. But if we do not have basic liberal rights to live and act with integrity, then what would ground our basic rights to freedom of religious belief and expression?

Martha Nussbaum argues that the source of our rights to religious freedom lies in our common ability to search for the ultimate meaning of life and for its intrinsic worth and value.⁴⁷ This searching faculty merits respect, and our rights to freedom of religion are rights that exist to protect this faculty, its exercise, and the expression of its exercise. She takes her account to be different from the integrity-based approach because, for instance, the judgment that one morally ought not go to war need not be the result of a search for the ultimate meaning of life and its intrinsic worth and value.⁴⁸

However, if Nussbaum's narrower meaning-, worth-, and value-based account of our rights to religious belief and expression holds, then several of the arguments I have made for the conclusion that trans people have rights to freedom of legal gender identification are still sound. For on Nussbaum's account, we have basic liberal rights to live a life that we take to be meaningful or valuable for us. And as I have been arguing, many trans people, after having engaged in a search for what a meaningful, worthwhile life for them involves, take such a life for them to involve their living as a gender different from the gender they were assigned at birth; and not being able to change their gender markers so that they do not clash with their gender identities clashes with their living such a life.⁴⁹ Put a different way, the arguments regarding authenticity, meaningful lives, and perhaps also misrepresentation that I have made for why trans people have basic liberal rights to freedom of legal gender identification still go through even if we accept (or should accept) Nussbaum's account of the grounds of our basic liberal rights to freedom of religious belief and expression rather than the integrity-based account.

47 Nussbaum, *Liberty of Conscience*, 168.

48 Nussbaum, *Liberty of Conscience*, 172.

49 See also Ashley, "What Is It Like to Have a Gender Identity?"

Alternatively, we might think that our basic liberal rights to freedom of religious belief and expression are grounded in our rights not to have to breach our perceived moral obligations—not to have to do something that we judge to be morally wrong. However, Nussbaum and Laborde plausibly argue that many religious believers do not take themselves to be morally obligated to practice their religion yet still seem to have rights to religious belief and expression.⁵⁰ In this case, their rights to religious belief and expression would have to be grounded in rights beyond rights to act in line with our perceived moral obligations. So we should reject a perceived-moral-obligation-based account of rights to freedom of religious belief and expression.

Second, the arguments that I have given might seem to show only that trans people who “pass” or are generally recognized as the gender that matches their gender identity have rights to freedom of legal gender identification.⁵¹ This might seem objectionable because it might seem that only a small minority of trans people “pass,” and trans people have no obligation to “pass.”

However, first, it is not obvious that only a small minority of (binary) trans people will be outed by identification documents that label them as genders that conflict with their gender identities. For instance, one study found that 28 percent of trans women generally “pass” as women and 62 percent of trans men generally “pass” as men.⁵² Furthermore, it is much easier to be accepted as the gender one presents as by a distracted and uninterested shop assistant or security guard who only briefly glances at you than it is to “pass” generally. (Many trans women do not “pass” because of how their voices sound or because of how their voices sound in prolonged conversations, but because of their appearance, they may nevertheless be assumed to be women by airport security, for instance.) And some trans people who do not exactly “pass” as the gender that matches their gender identity are people whose gender is not assumed or known by many people whom they encounter. Identification documents that label these trans people as particular genders lead to these trans people being thought to be particular genders (e.g., someone’s being thought to be a woman rather than unknown). And so identification documents that

50 Nussbaum, *Liberty of Conscience*, 172; and Laborde, *Liberalism’s Religion*, 66–67.

51 For problems with “passing” terminology (and suggestions for alternatives), see Serano, *Whipping Girl*, 176–80; and Plemons, *The Look of a Woman*, 14–15.

52 To et al., “Visual Conformity with Affirmed Gender or ‘Passing.’” This study was based on trans people’s perceptual reports. But many trans people are constantly worried about being perceived as (or expect to be perceived as) a gender different from their gender identity even when they are not so perceived. So we should not necessarily expect trans people to overestimate (rather than underestimate) the extent to which they “pass.”

clash with these trans people's gender identities will lead to them being treated as a gender that conflicts with their gender identity.

Of course, there are some trans people who are always thought to be the gender they were assigned at birth. However, first, identification documents that present a "nonpassing" trans woman as a man may still lead to her being treated as a man rather than as a woman by those with whom she interacts. For instance, such documents may lead to those to whom she presents her documents just thinking that she is a man and treating her as such rather than thinking that she is a trans woman and treating her as a woman. Finally, an important part of my argument is that trans people have *pro tanto* integrity-based rights not to have to present themselves to others as a gender that does not match their gender identity regardless of the extrinsic effects of this because many trans people judge that they ought not present themselves to others as a gender that conflicts with their gender identity, or judge that they would be misrepresenting themselves by doing this or that doing such is out of line with their practical identities. This part of the argument holds for all trans people regardless of whether they "pass" or are potentially outed by identification documents that do not match their gender identities.

A third objection to my argument for *pro tanto* rights to freedom of legal gender identification is that our passports and other identification documents present our sex, not our gender. However, first, as I discussed briefly at the start of this article, this is incorrect regarding many states and their identification documents. For instance, the Australian government notes that it collects, and its passports present, information about gender rather than sex—despite using the categories of male, female, and X.⁵³ Second, the laws of many countries draw an explicit link between sex and gender on legal identification documents even if they conceive of the marker on legal documents as marking sex. For instance, as mentioned earlier, Hungary forbids someone who was assigned female at birth from having a name from the state's list of names for men presented on their identification documents.⁵⁴ And the United Kingdom allows trans people to change whether they are presented as male or female on their passport so long as they have been diagnosed with gender dysphoria; it does not require that only biologically male people be listed as male, for instance. Thirdly and perhaps most importantly, the core argument that I have made goes through regardless of whether a legal document purports to present sex or gender information. For the argument that I have made is that, for instance,

53 See Australian Government, "Australian Government Guidelines on the Recognition of Sex and Gender," 4.

54 Andersson, "I Won't Even Be Allowed to Use My Name Now that Hungary Has Scrapped All Rights for Trans People."

a trans woman's being forced to have a passport that lists her as male will (1) force her to be outed as trans, (2) force her to present herself as a man, and/or (3) lead to others treating her as a man; and so it will impinge on her integrity.

2. ALL-THINGS-CONSIDERED RIGHTS TO FREEDOM OF LEGAL GENDER IDENTIFICATION

In the previous section, I established that trans and nonbinary people have *pro tanto* rights to freedom of legal gender identification. But these *pro tanto* rights could be outweighed by the rights of others or by harms to others. Relatedly, to establish that we have (*pro tanto*) rights to freedom of legal gender identification is not yet to establish what this means for the policies that we ought to have. In this section, I argue that (1) there are no rights, harms, or other considerations that outweigh trans and nonbinary people's *pro tanto* rights to freedom of legal gender identification. And I argue that because 1 is true, (2) the existing restrictions in many states on trans and nonbinary people's rights to freedom of legal gender identification are not justified and breach trans and nonbinary people's all-things-considered rights; and (3) there is a strong presumption in favor of one of several policies that make it relatively easy for trans and nonbinary people to have identification documents with gender markers that do not clash with their gender identities.

2.1. Blanket Bans on and Surgery Requirements for Gender Marker Change

First, I want to consider whether rights to freedom of legal gender identification generate all-thing-considered rights against blanket bans on gender marker change such as Hungary's, Kansas's, North Dakota's, Montana's, Oklahoma's, Florida's, Texas's, and Tennessee's blanket bans on gender marker change on birth certificates and driving licenses, and the US government's blanket ban on gender marker changes on passports. Do others' rights or the alleged harms averted by these blanket bans establish that these blanket bans do not violate anyone's all-things-considered rights? I consider these blanket bans on gender marker change simultaneously with another type of policy. Several US states, including Alabama, Arizona, Missouri, Nebraska, and Wisconsin (as well as many countries including Singapore and, until recently, Japan), require trans people to show proof that they have had sex reassignment surgery before they are permitted to change their gender markers on their legal documents.⁵⁵ Up to 90 percent of trans women have not had this kind of surgery, and up to 95

55 National Center for Trans Equality, "Summary of Birth Certificate Gender Change Laws." See also "Identity Document Laws and Policies" by the Movement Advancement Project (note 3 above).

percent of trans men have not have this kind of surgery; and many trans people do not want these expensive and invasive surgeries.⁵⁶ As I will argue in this section, it is implausible that alleged harms to others or the rights and interests of others can outweigh trans people's rights to freedom of legal gender identification and justify either a blanket ban or surgery requirements on trans people changing their gender markers.

Philosophers Holly Lawford-Smith and Kathleen Stock argue that we should not make it relatively easy for trans people to change their gender on their legal identification documents because this would put cis women at risk of harm in women-only spaces.⁵⁷ They seem to take their argument for this conclusion to favor policies like blanket bans.⁵⁸ This is their argument:

- P1. Cis men's being in women-only spaces would put cis women (and other people assigned female at birth) at risk of harm.
- P2. Many trans women share certain features of cis men that make them more likely to oppress and inflict violence on women—namely, high levels of testosterone, male genitals, and a history of having been socialized as men and treated as men. So permitting many trans women to use women-only spaces would put cis women (and other people assigned female at birth) at risk of harm.
- P3. If it were relatively easy to change one's gender on one's legal identification documents, cis men who wish to harm cis women (and other people assigned female at birth) could pretend to be women and thereby gain access to women-only spaces.
- C. Permitting trans women to change their gender on their identification documents relatively easily would harm cis women (and other people assigned female at birth) in women-only spaces because it would make it easier for cis men and trans women who pose a risk of harm to cis women in these spaces to access these spaces.⁵⁹

I do not want to evaluate the merits of this argument yet; I will do that in section 2.3 below. First I want to establish that this argument could not justify

⁵⁶ James et al., "The Report of the 2015 US Transgender Survey," 101–2.

⁵⁷ Lawford-Smith, *Gender Critical Feminism*, 104–5; and Stock, *Material Girls*, 106–8.

⁵⁸ Lawford-Smith advocates for blanket bans in her philosophical articles (e.g., "Ending Sex-Based Oppression"), as well as in her popular work.

⁵⁹ This line of argument is also extremely popular in popular culture and is generally cited as the reason that those who are in favor of strong restrictions on trans people changing their gender markers are in favor of these restrictions. See, e.g., Bland, "Wednesday Briefing." Lawford-Smith approvingly quotes a gloss of her argument along the lines of the reconstruction presented here (*Gender Critical Feminism*, 104–5).

blanket bans or surgery requirements on gender marker change. There are at least two reasons for this. First, if there were still significant barriers on changing one's gender markers on one's legal documents, cis men would not be able to change their gender markers on their legal documents in order to access women-only spaces without bearing costs that no cis man would be willing to bear to do this. A state can allow trans women to change their legal gender markers but require that before doing this, they show documented evidence that they have lived as a woman for several months or years or that they have been on hormone replacement therapy for several months or years.⁶⁰ These are extraordinarily high costs to bear for a cis man. And these costs will not be borne by cis men who wish to access women-only spaces, especially since most cis men who wish to access these spaces simply walk into them.⁶¹

Second, many trans women do not have the features alluded to in P2. Around 70 percent of trans women have undertaken feminizing hormone replacement therapy (HRT); and 95 percent of trans women want to be on feminizing HRT.⁶² Feminizing HRT lowers one's testosterone levels to average cis women levels and raises one's estrogen levels to average cis women levels.⁶³ There is no reason to believe that someone's having male genitals on its own—that is, without testosterone levels higher than average cis women levels and estrogen levels lower than average cis women levels—has any connection to harming cis women. And some trans women transition very young and so have not been socialized as men (and have been on feminizing HRT since transitioning). Lawford-Smith's and Stock's argument could not justify restricting the integrity-based rights of such trans women. So blanket bans and surgery requirements on gender marker change are unjust because they encroach on the integrity-based rights of many trans people without justification.

I have been asked: Why would it be better to (1) require that trans people have had HRT before being able to change their gender markers rather than to (2) require that they have had sex reassignment surgery before doing this? To be clear, I am not arguing that either 1 or 2 are justified requirements. I am arguing that Lawford-Smith's and Stock's argument cannot justify 2 over 1. However, 1 is better than 2 because over 90 percent of trans men and trans women have not had sex reassignment surgery, and this surgery can be extremely costly and invasive. HRT is not so costly nor so invasive, and most trans women, for

60 See further section 2.3 below.

61 See Doran, "Equality NC Director"; and Steinmetz, "Why LGBT Advocates Say Bathroom 'Predators' Argument Is a Red Herring."

62 James et al., "The Report of the 2015 US Transgender Survey," 99.

63 Vincent, *Transgender Health*, 152.

instance, are either on HRT or would like to be. So 1 encroaches less strongly on trans people's integrity-based rights than 2.

It has been put to me that my argument that surgery requirements and blanket bans on gender marker change cannot be justified by P1-C does not succeed because we need generalized policies regarding gender marker change, and I have not shown that we can hold such a generalized policy without adopting a blanket ban or surgery requirements. This is not the case. There are several general policies that I have argued that the argument from P1-C does not militate against and are inconsistent with blanket bans on gender marker change and surgery requirements, such as the following: trans people who transition before they are fourteen years old may change their gender markers on their identification documents; trans people who have been on HRT or who have lived as the gender that matches their gender identity for, e.g., three to six months (or twenty-four months) may change their gender markers. (A three-to-six-month policy regarding legal gender change was proposed by the Scottish government; a twenty-four-month period is the current UK policy regarding birth certificates.)⁶⁴ To reiterate, I am not arguing that any of these gatekeeping policies can be justified. I am just arguing that Lawford-Smith's and Stock's arguments do not justify blanket bans on gender marker change rather than these gatekeeping policies.

2.2. *No X Markers on Passports*

In the United Kingdom there is no option to have an X marker, which denotes neither binary gender, on one's passport.⁶⁵ After allowing X markers on passports from 2021, in early 2025, the US federal government returned to a policy of forbidding X markers on passports.⁶⁶ Forbidding X markers on passports encroaches on nonbinary people's ability to live and act with integrity (as outlined in section 1.2 above). And allowing nonbinary people to change their gender marker to X would not lead to any of the harms that Lawford-Smith and Stock are concerned about that might outweigh nonbinary people's integrity-based rights to have X markers on their passports. This is because having an X marker on one's documents or being legally nonbinary does not enable one to access women-only spaces.

The United Kingdom and United States could revise their laws in order to preserve the integrity of nonbinary people in several ways. They could allow X

64 On the Scottish proposal, see section 2.3 below; on the UK policy, see HM Passport Office, "Guidance."

65 See HM Passport Office, "Guidance."

66 See note 4 above.

markers on passports. But another option would be to entirely remove gender markers from passports. Some academic lawyers have argued that the state should decertify gender. If gender is decertified by a state, then that state stops collecting and presenting information about gender. If the United Kingdom decertified gender, it would treat its citizens' genders in a way that is similar to how it currently treats its citizens' religious identities and disabilities: the UK state does not collect and present information about religious identities and disabilities except in the census.⁶⁷ Rather than fully decertifying gender, the United Kingdom could also simply stop presenting gender information on passports or stop requiring that this information be presented on passports. The Australian state of Tasmania has adopted a similar policy with birth certificates: gender is no longer mandatorily listed on birth certificates issued in Tasmania, though parents can opt in to listing a gender on a child's birth certificate.⁶⁸ In order to preserve nonbinary people's integrity, the United Kingdom and United States could allow X markers, decertify gender, or adopt a Tasmania-style opt-in gender policy for UK and US passports. But the UK's and US's current policies that require binary gender markers on all UK and US passports is an unjust encroachment on nonbinary people's integrity.

2.3. *A Presumption in Favor of Self-Identification, Decertification, or Similar Policies*

There are several laws and policies that have been proposed and, in some places, implemented that (would) make it relatively easy for trans people to ensure that their gender markers on their legal identification documents do not clash with their gender identities. *Self-identification* policies allow trans people to change their gender markers by simply declaring themselves to be a gender other than the gender they were assigned at birth and, at most, paying the administrative fee needed to cover the costs of changing their gender markers on identification documents such as their driver's licenses and passports. Over thirty countries have self-identification policies, including Argentina, Brazil, Ireland, Norway, Portugal, and Spain.⁶⁹

Decertification policies ensure that trans people do not have identification documents that clash with their gender identities by not presenting gender markers on anyone's legal identification documents. Related to generalized decertification are policies that stop the presentation of gender information on particular legal documents: for instance, although US states present gender

67 Renz and Cooper, "Reimagining Gender Through Equality Law."

68 Gogarty, "All Colours of the Rainbow."

69 ILGA Europe, "Annual Review of the Human Rights Situation of Lesbian, Gay, Bisexual, Trans and Intersex People in Europe and Central Asia"; and Chiam et al., "Trans Legal Mapping Report 2019."

information on driver's licenses, Australian states and the United Kingdom do not. A further related policy would be a Tasmania-style policy that makes the presentation of gender information on passports optional.

Finally, there are other policies that make it relatively easy for trans people to change their gender markers on their legal documents. For instance, in late 2022, the Scottish Parliament passed a bill allowing all trans people to change their legal gender and gender markers so long as they demonstrate that they have lived as that gender for three to six months. We can call this the *Scottish proposal*. In early 2023, the UK Parliament vetoed this bill.⁷⁰

As I will explain, if we have integrity-based rights to freedom of legal gender identification, there is a strong presumption in favor of enacting self-identification, decertification, or the Scottish proposal. Trans people's basic liberal rights to live and act with integrity require that they be able to change their gender markers on their legal documents. And the bar for restricting our basic liberal rights to live and act with integrity is high: we may restrict these rights only if we have very good reason to believe that refraining from restricting these rights will infringe on others' basic liberal rights or otherwise seriously harm them. But the arguments for restricting trans people's rights and not enacting one of self-identification, decertification, or the Scottish proposal simply do not have this high evidential caliber.

Several arguments have been made against these policies that make it relatively easy for trans people to ensure that their gender markers do not clash with their gender identities. But these arguments fall into two categories: they are either *irrelevant* or *implausible*. First, the irrelevant arguments concern the negative effects of these policies for cis women in women's sports, prisons, and domestic violence shelters, and for trans children being prescribed supposedly dangerous puberty blockers. Such arguments have been made by Stock and Lawford-Smith.⁷¹ *Contra* Stock's and Lawford-Smith's arguments, these policies regarding gender markers simply have no implications regarding these other domains. For instance, many trans women are legally recognized as women and have gender markers that match their gender identity on all of their legal documents, but all trans women are currently banned from participating in many international and domestic sporting competitions: World Rugby, World Aquatics, and World Athletics have recently banned all trans women, or all trans women who have gone through any stage of male puberty, from participating in their competitions—and this includes trans women who are legally recognized as women. And British Triathlon and England's Rugby

70 Bland, "Wednesday Briefing."

71 Stock, *Material Girls*, 105–18, 83–89; and Lawford-Smith, *Gender Critical Feminism*, 95–111.

Football Union and Rugby Football League are among domestic sporting organizations that have banned all trans women, including those legally recognized as women, from participating in their competitions.⁷² Some intersex women have also been legally recognized as women for their entire lives but are unable to participate in women's sports.⁷³ So policies regarding trans women's gender markers have no implications for trans women's eligibility to participate in women's sports.

Similarly, if providing puberty blockers to trans children did genuinely harm them, this could justify not providing puberty blockers to trans children while enabling them to change their gender markers so that they do not clash with their gender identities.⁷⁴ Indeed, the UK has recently adopted a policy along these lines.⁷⁵

In the UK, domestic violence shelters have an exemption to the Equality Act that allows them to refuse to admit trans women, including trans women who are legally women and have gender markers that match their gender identity on all their legal documents.⁷⁶

If trans women were genuinely a risk to cis women in women's prisons, this risk could justify segregating trans women from cis women consistent with trans women's being legally recognized as women. Regardless, the current UK policy also requires most trans women prisoners who are legally women to be housed in men's prisons.⁷⁷ So it is clear that arguments regarding the negative effects in different domains of making it relatively easy for trans people to have gender markers on their legal identification documents that do not clash with their gender identities are irrelevant.

It might be objected that if freedom of legal gender identification really did not have implications for gender-affirming health care, prisons, sports, and domestic violence shelters, then this would show that freedom of legal gender identification is not that important or does not get trans people that much. But this is not so. As explained in section 1 above, trans people face significant costs

72 See BBC, "Fina Bars Transgender Swimmers from Women's Elite Events" and "UK Athletics Wants Open Category for Male and Transgender Athletes"; Reuters, "England's Rugby Union and Rugby League Ban Transgender Players from Women's Game"; and Roan, "British Triathlon Becomes First UK Sport to Create 'Open' Category for Transgender Athletes."

73 See, e.g., Savulescu, "Ten Ethical Flaws in the Caster Semenya Decision on Intersex in Sport."

74 Alternatively, permissions to change one's gender markers could be restricted to adults.

75 Trigg, "Puberty Blockers for Under-18s Banned Indefinitely."

76 Equality and Human Rights Commission, "Separate and Single-Sex Service Providers."

77 Ministry of Justice et al., "New Transgender Prisoner Policy Comes into Force."

to their ability to live with integrity and significant risk of harm if they are forced to have legal identification documents that out them as trans and/or label them as genders that clash with their gender identities. We must use legal identification documents all the time to access a variety of goods and services. Granting freedom of legal gender identification therefore (in itself) has significant beneficial consequences for trans people. And a state that forces trans people to have legal identification documents that clash with their gender identities constantly breaches trans people's significant rights and subject trans people to significant harm. So freedom of legal gender identification is important on its own.

An argument that may not be irrelevant concerns the effects of allowing trans women to use women's restrooms. I say that this argument *may not be irrelevant* because several US states, including Florida, have or have had bathroom bans that prohibit trans women—including trans women who are legally recognized as women and have gender markers that match their gender identities—from using women's restrooms.⁷⁸ In early 2024, members of the UK government proposed a similar policy, and in 2025, the UK government's Equality and Human Rights Commission issued guidance according to which trans women are not permitted to use any public or private women's restrooms (including trans women with female gender markers on all of their identification documents).⁷⁹ So arguments concerning the alleged harms to cis women in restrooms may not be relevant to questions about the policies regarding gender markers that justice requires that we have because trans women with female gender markers on their identification documents could still be banned from women's restrooms. But the argument that making it relatively easy for trans people to change their gender markers would harm cis women in restrooms is also *implausible*. As noted above, over thirty countries have self-identification policies, and there is no evidence from these countries that such policies have resulted in a spike in harm to cis women in women's restrooms.⁸⁰ So there is no evidence that

78 Ables, "Florida Passes Bathroom Bill in Latest Wave of Anti-Trans Legislation."

79 Elgot, "Kemi Badenoch Could Rewrite Law to Allow Trans Exclusion from Single-Sex Spaces." See also Stavrou, "Trans Women to Be Banned from Single-Sex Spaces Under New EHRC Guidance."

80 See Middleton, "Scotland's Trans Self-ID Bill No Risk for Women, Says UN Expert"; and Kelleher, "Ireland Has Had Trans Self-ID Laws for Years." Opponents of gender self-identification in Ireland do not cite any harms to cis women in restrooms resulting from Ireland's self-identification policy in the years since it was implemented in 2015; the only evidence they give of the policy's harms involve a lack of consultation with the Irish public regarding implementation and the fact that one violent trans woman may be housed in a women's prison in Ireland as a result of Ireland's particular policies regarding gender. See, e.g., the post by ripx4nutmeg, "Ireland Has Had Self ID for Years, and There Haven't Been Any

self-identification policies harm cis women in restrooms.⁸¹ Proponents of the claim that self-identification policies result in harm to cis women in restrooms cite no evidence to support this claim.⁸² Furthermore, there is growing evidence that one of the main results of banning trans women from women's restrooms is that cis women begin to be harassed by other cis women who mistakenly believe them to be trans.⁸³ So such bans in fact seem to harm cis women.

Perhaps the argument from harms in restrooms against self-identification, decertification, and the Scottish proposal will at some stage be made watertight and made stringently. But at the moment it does not seem that this argument has the evidential caliber to justify restricting basic liberal rights to live and act with integrity.

Proponents of arguments against policies like self-identification, decertification, and the Scottish proposal sometimes suggest that proponents of these

Problems," *Glinner Update* (blog), May 11, 2022, <https://grahamlinehan.substack.com/p/ireland-has-had-self-id-for-years>; and Hayton, "How the Trans Activists Fooled Ireland."

- 81 See also Steinmetz, "Why LGBT Advocates Say Bathroom 'Predators' Argument Is a Red Herring"; and Serano, "Transgender People, Bathrooms, and Sexual Predators." It might be argued that although there is no evidence from these countries of a spike in harms in women's restrooms, this is just because it would be difficult or offensive to collect this data. However, this is not the case. It would not be difficult or offensive to collect evidence of a rise in harm (or lack thereof) to women in women's restrooms over a particular time-frame that coincided with the adoption of a self-identification policy. Indeed, Hasenbush et al. precisely assessed whether privacy and safety concerns and violations in women's restrooms had increased in US localities that had adopted ordinances permitting trans women to use women's restrooms and locker rooms; they found no evidence that they had ("Gender Identity Nondiscrimination Laws in Public Accommodations," 78).
- 82 The most that Stock does to support this argument in her published work is to provide one case of a trans woman who assaulted a cis woman in a public toilet (*Material Girls*, 106). But this is just one case; there are also cases of cis women attacking and harassing both other cis women and trans women in public toilets. See Halberstam, *Female Masculinity*, 19; Warr, "How Do Gender Non-Conforming Individuals Experience Gendered Public Toilets?"; Billson, "Cis Woman Harassed by Transphobe' Who Followed Her into Female Toilet Because She Has Short Hair"; Lopez, "Women Are Getting Harassed in Bathrooms Because of Anti-Transgender Hysteria"; Brooks, "I've Been Spat On"; and Stewart, "2 Women Charged with Sexual Battery of Trans Woman in North Carolina Bar." And the fact that one trans woman assaulted a cis woman in a women's public toilet cannot justify stopping all trans women from using women's public toilets. Similarly, the fact that one neonatal nurse who was a British cis woman killed seven infants in her care cannot justify forbidding all British cis women from being neonatal nurses.
- 83 Halberstam, *Female Masculinity*, 19; Warr, "How Do Gender Non-Conforming Individuals Experience Gendered Public Toilets?"; Billson, "Cis Woman Harassed by Transphobe' Who Followed Her into Female Toilet Because She Has Short Hair"; Lopez, "Women Are Getting Harassed in Bathrooms Because of Anti-Transgender Hysteria"; and Brooks, "I've Been Spat On."

policies must show that they would not harm cis women. As I have explained, there is a good case that proponents of these policies have shown this—that is, have shown that allowing trans people to (relatively) easily change their gender markers so that they do not clash with their gender identities does not harm cis women, or at least that we have no reason to believe this. Regardless, if trans people have integrity-based rights to freedom of legal gender identification, the burden of proof does not lie with proponents of these policies; it in fact lies with those who argue against them. For it must be shown that there is very strong evidence that such policies will harm cis women in order for encroachments on trans people's integrity-based rights to be justified. And this has not been shown.

2.4. *Objections*

I have encountered several objections to my argument that trans people's all-things-considered rights establish that we should adopt a self-identification policy, decertification, or the Scottish proposal. First, a referee has put it to me that there is a disanalogy between (1) integrity-based rights to religious exemptions and rights for ethnic Malays to be able to not have 'Muslim' presented on their identification documents and to not be treated as Muslims by the Malaysian state and (2) integrity-based rights not to have the gender on one's identification documents present one as a gender that conflicts with one's gender identity. The referee argues that regarding 1, there is never a question of having to participate or practice another's religion or conception of the good in order to satisfy their integrity interest, but in the case of 2, this is not so. The referee mentions several cases in which trans people's gender marker change on their legal documents gives rise to obligations for others to believe things or act as if they believe them or face a discrimination allegation. First is the Australian case of *Tickle vs. Giggle*, in which the founder of a social media app was found to have unlawfully discriminated against trans woman Roxanne Tickle—who is legally a woman because she has changed her birth certificate to present her as female—by excluding Tickle from her women-only social media app. Second is the case of an Australian trans woman who is similarly legally a woman and so cannot be excluded from a woman's hockey team. I take it that this objection is best understood as an objection to the idea that trans people have *all-things-considered* rights to not have gender markers that conflict with their gender identities. The argument here is not that these cases and this contrast between 1 and 2 show that trans people have no such integrity-based *pro tanto* rights to change their gender markers but only that considerations about the knock-on effects of granting these rights for what others must do can make it the case that trans people do not have all-things-considered integrity-based rights to easily have gender markers that do not clash with their gender identities.

There are several things to say in response. First, there is not a contrast between the Malaysian case and the trans case here. Being assigned Muslim at birth in Malaysia involves being expected to be an observant Muslim, to engage in Islamic fasting and prayer, and to be subject to fines for failing to do these things.⁸⁴ Changing one's Malaysian identification documents so that they no longer label one as Muslim implies that the state and others may not hold one to these obligations and so may not treat one as Muslim. Second, I have argued only that trans people have integrity-based rights *not to be presented* as a gender that conflicts with their gender identity on their legal documents. One way for a state to satisfy this negative duty is for it not to present gender information or gender markers on any documents (i.e., decertification). If a state does not present any gender information on its legal identification documents, the information so presented cannot yield obligations for others, for there is no such information. Third, we must distinguish between

1. gender and gender marker change on legal identification documents; and
2. gender and gender marker change on legal identification documents having implications for the gender that others must treat someone as on pain of unlawful discrimination.

I have been making an argument only regarding 1 in this paper. Nothing I have said bears on whether it should be unlawful discrimination to exclude trans women from women-only spaces or women-only apps. Those who think that it should be permissible to do this can accept everything that I have said but hold that it is permissible for, for instance, those with different views about the metaphysics of gender to exclude trans women from women-only spaces. As I explained in section 2.3, 1 need not, and in many places does not, imply anything regarding 2.

But what of the referee's two cases? First is the question of whether Australian trans women who are legally women can be lawfully excluded from a women's hockey team. As I explained in section 2.3 above, all trans women, including those who are legally women, are excluded from many professional women's sports. Trans women are also excluded from women's sports in Australia. For instance, Basketball Australia forbade trans woman Lexi Rodgers from playing for a women's basketball team in 2023 and did not face any allegations of unlawful discrimination.⁸⁵ So it is not true that Australian women's sports

84 See the references in note 20 above.

85 See Guardian Sport, "Basketball Australia Rules Transgender Athlete Lexi Rodgers Ineligible to Play."

teams must allow trans women who are legally women to participate in their teams or face a discrimination lawsuit. In the second case, the Australian Federal Court found that it was unlawful discrimination for a trans woman who is a woman according to her birth certificate to be excluded from a women-only social media app in *Tickle vs. Giggle*. But although such exclusion is unlawful discrimination in Australia, it is not in other places. For instance, in 2024, the UK government stated that trans women—including trans women who are legally women because they have gender recognition certificates and have changed their birth certificates to present them as women—do not have legal rights to access women-only spaces.⁸⁶

So my argument is about only (1) whether, for instance, trans women have all-things-considered rights not to be labeled as men on all their legal identification documents, not (2) whether trans women must always be treated as women and not excluded from women-only spaces. These two issues are separable. The arguments for 1 do not establish 2 on their own, and we must be careful to distinguish what our arguments are for and what they support: arguing that 2 should be permissible (that it should be permissible to exclude trans women from women-only spaces and apps) does not establish that it should be permissible for a state to label trans women as men.⁸⁷

Finally, if it were somehow impossible to distinguish 1 and 2 in a particular jurisdiction (and if decertifying gender were infeasible), and it were concluded that people or service providers should be able to exclude trans women from women-only spaces, then one response would be to allow freedom of legal gender identification to the greatest extent possible without creating changes that force people and service providers not to exclude trans women from women-only spaces. For instance, first, in jurisdictions where one's legal gender is determined by the gender on one's birth certificate, gender change by self-identification could be permitted on all legal documents except birth certificates. Some, such as some young or poor people, have no identification documents other than

86 Office for Equality and Opportunity, "Response to Call for Input on Single-Sex Spaces." See also the other cases discussed in section 2.3 above. It should be noted that in 2025, legal decisions in the United Kingdom have been made that may be understood to imply that no trans woman is legally a woman; but in 2024, at least trans women with birth certificates presenting them as women were taken to be legally women.

87 It should be noted that *Tickle vs. Giggle* does not establish that gender-critical feminists cannot, for instance, exclude trans women from their events. It just establishes that spaces and apps that label themselves women-only cannot exclude trans women who are legally women on the basis that they are not women. One may still hold an event for cis women, people assigned female at birth, or even "biological females," consistent with this; and (Australian) gender-critical feminists do this. Similarly, trans organizations have spaces and events for only trans people and for only trans women.

birth certificates, so there are strong reasons not to adopt this kind of policy.⁸⁸ However, this policy better preserves trans people's integrity-based rights than many US states' policies, such as those in Kansas, Montana, North Dakota, Tennessee, Oklahoma, Florida, and Texas, as well as the US federal government's policies and the UK government's policies (since one cannot change one's gender on one's UK passport through self-identification). A second alternative policy is inspired by the Australian state of Tasmania, which has made gender information opt-in on birth certificates while still collecting sex/gender information for children born in Tasmania; this information is simply not presented (by default) on birth certificates. Similarly, a state might collect legal gender information about its citizens in some way that is not determined by self-identification, but although it may collect this information, it may not present it (or not present it by default) on any legal identification documents. The all-things-considered rights articulated in this paper seem consistent with such a policy.

A referee has put a different objection to me regarding my argument that trans people's all-things-considered rights establish that we should adopt self-identification, decertification, or the Scottish proposal. This referee argues that there are other arguments against a self-identification policy that do not concern physical harm, such as (1) "harms to women and the project of sex equality [as a result] of the new understanding of woman suggested by a self-identification policy ('a woman is anyone who identifies as a woman')" and (2) "the expressions of those born male who declare themselves as women [which] can strengthen stereotypes about what it means to be a woman or female."

However, regarding 1, I have not argued that we or our state should adopt the view that a woman is anyone who self-identifies as a woman. I have argued only that we should adopt either self-identification, decertification, or the Scottish proposal as policies regarding gender markers on identification policies. Those concerned about our state legally defining women as those who self-identify as women might opt for decertification, which involves the state not presenting (or not collecting and presenting) information about people's genders and so not adopting a legal definition of sex or gender.⁸⁹ Alternatively,

88 Neuman Wipfler, "Identity Crisis," 537.

89 The UK does this for religion, race, and disability. For discussion, see Renz and Cooper, "Reimagining Gender Through Equality Law." Alternatively, a self-identification policy for gender on legal documents might be undertaken, but the state might not take self-identification to be what uniquely makes someone a woman or socially treated as a woman. The state might take up a pluralist of what it is to legally be a woman, for instance. It might take self-identifying as a woman to be sufficient for being a woman (in terms of one's legal documents) but not necessary, since those who are assigned female at birth and do not change their gender markers do not necessarily self-identify as women or have legal documents that are sensitive to any such self-identification. At the same time, our state (or its

they might opt for (or argue for) the Scottish proposal, which does not define a woman as anyone who self-identifies as a woman but rather holds that to legally be a woman, one must have been assigned female at birth or lived as a woman for at least three to six months.⁹⁰

Regarding 2, the variety of gender expressions of trans women is as large and as varied as the variety of gender expressions of cis women. So adopting self-identification would not strengthen the view that women ought to look a certain way. But if it somehow did, again, decertification would be a good alternative, since the removal of gender markers from legal documents does not involve claiming that trans women are women because they express themselves in particular ways; it does not involve holding that trans women are—or anyone is—any particular gender.⁹¹

So it seems that these objections do not undermine my case that trans people's all-things-considered rights establish that we should adopt self-identification, decertification, or the Scottish proposal.

3. ALTERNATIVE APPROACHES

There is little existing philosophical work on the moral grounds of trans rights to freedom of legal gender identification. But there are some brief discussions of the grounds of trans rights in general, there are some discussions of other trans rights in bioethics, and there is discussion in the law literature of the legal grounds of trans people's rights. These discussions present alternative pictures of the grounds of trans rights. In this section, I will argue that integrity provides

legal system) might hold that to suffer sex- or gender-based discrimination, one need not self-identify as or be the relevant gender on one's birth certificate. For instance, someone who is hired as a woman but who subsequently self-identifies as nonbinary (and changes their birth certificate and other legal documents accordingly) might still be treated as a woman and might be subject to discrimination at work or elsewhere on this basis. It is consistent with a self-identification policy regarding identification documents that we hold that this nonbinary person is subject to sex/gender discrimination because they are treated as a woman—which has nothing to do with their gender on legal documents.

90 It is unclear why someone worried by the idea that a self-identification approach would define what it is to be a woman in terms of self-identification as a woman would prefer a law that says that one can be legally a woman so long as one has lived as a woman for two years (as per UK law from 2004 to 2024) over a law that says that one can be legally a woman so long as one has lived as a woman for three to six months (as per the Scottish proposal).

91 Some might worry about how gender discrimination and single-sex spaces would work under decertification. For sustained discussion of this issue and of how this might and can work, see Grabham, "Decertifying Gender"; Renz, "Gender-Based Violence Without a Legal Gender"; and Renz and Cooper, "Reimagining Gender Through Equality Law."

a better account of the grounds of rights to freedom of legal gender identification than these alternatives.

3.1. *Autonomy*

Some legal theorists have taken legal rights to freedom of legal gender identification to be autonomy-based rights.⁹² In public, activist, and some academic discussions, many trans rights are taken to be autonomy based, and in bioethics, other trans rights have been argued to be autonomy based.⁹³ Furthermore, in one of the few discussions linking the potential moral grounds of rights to have the gender on our legal identification documents not clash with our gender identity to the legal grounds of these rights, Holning Lau argues that our personal autonomy provides a ground of this right.⁹⁴

However, many of these discussions do not fully explain what the relevant notion of autonomy is or why we should think that our autonomy generates moral rights to freedom of legal gender identification.⁹⁵ Some accounts of autonomy understand it to be very similar to the concept of integrity that I outlined in section 1.1. For instance, according to Ben Colburn, autonomy involves living in accordance with one's own conception of what is valuable.⁹⁶ And many other philosophers hold similar views to Colburn's, on which trans and nonbinary people's living with autonomy involves their living with integrity, in the sense discussed in this article.⁹⁷ To the extent that we should understand autonomy in the way that Colburn and these other philosophers do, the integrity-based account of rights to freedom of legal gender identification is an integrity/autonomy-based account of rights to freedom of legal gender identification. If we should

92 See Cannoot and Decoster, "The Abolition of Sex/Gender Registration in the Age of Gender Self-Determination," 32–33, 35, 42. See also Renz, "Genders that Don't Matter," 11; and Ashley, "'X' Why?" 43.

93 See, e.g., Pearce et al., "Introduction," 15; Gerritse et al., "Decision-Making Approaches in Transgender Healthcare"; and Ashley, "Adolescent Medical Transition Is Ethical."

94 Lau, "Gender Recognition as a Human Right," 194–95.

95 See Cannoot and Decoster, "The Abolition of Sex/Gender Registration in the Age of Gender Self-Determination," 32–33, 35, 42. See also Renz, "Genders that Don't Matter," 11; and Ashley, "'X' Why?" 43.

96 Colburn, "Autonomy and Adaptive Preferences," 62.

97 Suzy Killmister distinguishes autonomy and integrity (*Taking the Measure of Autonomy*, 10), but similarly, on Killmister's account, acting out of line with one's judgments of how one ought to live and act involves both acting without integrity and (at least other things equal) acting without autonomy. Somewhat similarly, many understand the Rawlsian approach to basic liberal rights, which we discussed in section 1.1 and understood as integrity based, to be an autonomy-based approach to these rights. See, e.g., Christman, "Autonomy in Moral and Political Philosophy," secs. 3.1, 3.5.

understand autonomy (or one sense of it) as very similar to integrity, then we should understand my argument that we have rights to freedom of legal gender identification to be an argument that is in line with this idea that such rights are autonomy based. But since the case that there are such autonomy-based moral rights to freedom of legal gender identification has not been articulated in detail in existing literature, we should see my account as providing the first thorough articulation, development, and argument for the view that there are such autonomy-based rights to freedom of legal gender identification.

However, it is unclear that we should accept an account of autonomy like Colburn's that understands autonomy and integrity to be very similar. This is because other plausible conceptions of autonomy are broader than Colburn's and broader than the concept of integrity outlined in section 1.1. For instance, Stephanie Kapusta holds that an agent possesses personal autonomy *if they act on motives that are their own*.⁹⁸ As Sarah Buss and John Christman discuss, a view along these lines on which our acting autonomously is determined by our acting in line with (reflectively endorsed) motives is very much a, or even *the*, standard account of autonomy in moral and political philosophy.⁹⁹ But it is not clearly plausible that rights to freedom of legal gender identification could be grounded in our rights to be autonomous in this sense. For the idea that we have rights to be autonomous in this sense that can ground rights to freedom of legal gender identification would overgenerate rights. For instance, I might be strongly motivated to have my hobbies listed on my identification documents or to not have my age listed on my identification documents. But it does not seem to follow that I have a right to have my hobbies listed on these documents or to not have my age listed on these documents. So we should hold that there are integrity-based rights to freedom of legal gender identification, but it is unclear that it follows from this that we have autonomy-based rights to freedom of legal gender identification.

Is there another conception of autonomy or autonomy-based rights that might more plausibly ground rights to freedom of legal gender identification? This is not clear. For instance, in a forthcoming article, E. M. Hernandez and Rowan Bell briefly claim that trans rights are in general grounded in autonomy rights. They say a little about what they have in mind by 'autonomy rights'.¹⁰⁰ They explain that they endorse Thomas Hill's view that one's autonomy rights are one's rights "to make otherwise morally permissible decisions about matters

98 Kapusta, "Gender Autonomy," 346.

99 Buss and Westlund, "Personal Autonomy," esp. sec. 2; and Christman, "Autonomy in Moral and Political Philosophy," esp. sec. 1.

100 Hernandez and Bell, "Much Ado About Nothing," sec. 3.

deeply affecting one's own life without interference by controlling threats and bribes, manipulations and willful distortion of relevant information."¹⁰¹ However, if we strongly, deeply, and persistently desire something, then whether this strong desire is satisfied will deeply affect our own life. Yet as I have argued, we should not think that our strong desires regarding our identification documents on their own generate rights regarding them. For we might strongly desire to have our age not listed on these documents or to have our hobbies listed on them, but this does not show that we have rights to these things.

3.2. *Privacy*

A different grounding of rights to freedom of legal gender identification that has been to some extent discussed in the legal literature is in our rights to privacy.¹⁰² And in many jurisdictions, trans people have rights not to be forced to disclose the fact that they are trans and rights that others do not disclose this fact about them without their consent.¹⁰³ So we might think that one ground of our rights to change our gender markers on our legal documents so that they do not out us as trans lies in our privacy rights.

Our privacy might well provide us with rights here. However, privacy-based concerns may not be able to distinguish between (1) forcing a trans person to out themselves as trans by forcing them to have legal documents that present them as the gender they were assigned at birth and (2) forcing someone to disclose the city they were born in by forcing them to list it on their identification documents. Our rights to live with integrity provide a good account of why 1 is worse than 2. Furthermore, a privacy-based argument would seem to struggle to generate nonbinary rights to have X markers on identification documents. For having an X marker on one's identification documents is likely to out one as trans.

3.3. *Harm*

At the start of this article, I discussed the risk of harm that trans people are subject to if they are forced to present themselves as a gender that does not

101 Hill, *Autonomy and Self-Respect*, 48. Similarly, Lau characterizes our personal autonomy as our "freedom to make decisions about oneself, for oneself" ("Gender Recognition as a Human Right," 194). And Lau contrasts our gender identity with other aspects of our identity and other decisions that may not affect our life to the same degree (195).

102 See Cannoot and Decoster, "The Abolition of Sex/Gender Registration in the Age of Gender Self-Determination," 33–34; Lau, "Gender Recognition as a Human Right," 196–97; Weiss, "The Gender Caste System," 133, 168–73; and Hines, *Gender Diversity, Recognition and Citizenship*, 43. Cf. the Yogyakarta Principles (note 14 above) 14. Heath Fogg Davis notes that the case for removing race markers from identification documents was partially made on the basis of rights to privacy (*Beyond Trans*, 37).

103 See, e.g., Galop, "Trans Privacy Law."

match their gender identity and the gender that people assume them to be. Such harms are discussed by several lawyers and legal theorists as reasons that decertifying gender or removing gender markers from (some) legal documents would be a good idea—though not as reasons why we have rights to freedom of legal gender identification.¹⁰⁴ But perhaps trans rights to freedom of legal gender identification should be understood to be grounded in our rights not to be subject to such harms rather than grounded in integrity?

However, there are limits to what such a potential harm-based grounding can do. For instance, nonbinary people's being enabled to present themselves as nonbinary by being enabled to use X markers on their legal documents may subject them to more harm than they would otherwise be subject to. This is because nonbinary people are subject to abuse and assault, and without presenting themselves using identification documents that feature an X marker, nonbinary people would not (necessarily) present themselves to others as nonbinary; without presenting themselves with an X marker, most nonbinary people will be assumed to be a binary gender.¹⁰⁵ So a harm-based account struggles to generate nonbinary rights to X markers. But as I have argued, their rights to live with integrity provides nonbinary people with rights to X markers (if there are gender markers on legal identification documents).

Furthermore, a harm-based argument would not seem to provide rights to freedom of legal gender identification in a society in which transphobia has been eliminated; but the integrity-based argument that I made in sections 1 and 2 seems to show that even in a society that is not transphobic, where trans people are not subject to harm *qua* trans people, trans people have rights to change their gender markers so that those markers do not clash with their gender identities.

Finally, we might worry that such a harm-based account of trans rights to freedom of legal gender identification is too contingent upon what the right account of harm is. For instance, most accounts of harm hold that we are harmed whenever the state makes us do something that makes us very unhappy or that we prefer not to do.¹⁰⁶ But in this case, we do not have a *pro tanto* right that the state does not subject us to harm. For we might prefer to not present our date of birth or birth city on our identification documents (or be very unhappy if we are made to have this information on these documents), but this

104 See Neuman Wipfler, "Identity Crisis," 493, 536; Cooper and Emerton, "Pulling the Thread of Decertification," 7; Lau, "Gender Recognition as a Human Right," 197; Weiss, "The Gender Caste System," 173; and Ashley, "'X' Why?" 37.

105 See James et al., "The Report of the 2015 US Transgender Survey," 89; and Rajunov and Duane, *Nonbinary*, xxiv–xxvi.

106 See, e.g., Fletcher, *The Philosophy of Well-Being*.

preference or unhappiness does not seem to on its own establish that we have a right that our documents do not present this information about us. Similarly, we might strongly prefer to have our hobbies listed on our legal documents, but we do not have even a *pro tanto* right that the state allow us to do this. Harm may play some role in the case for trans rights to freedom of gender, but there is a good case that our rights to live with integrity also play a role and on their own are sufficient to ground strong rights to freedom of legal gender identification.

3.4. *Issues with Integrity*

I have been arguing that there are reasons to accept an integrity-based account of the grounds of rights to freedom of legal gender identification rather than alternative accounts. But there is an issue about the boundaries of integrity-based rights that might lead us to worry that we should favor an alternative grounding of rights to freedom of legal gender identification. We might wonder: Do pro-life pharmacists have integrity-based rights to refrain from selling the morning-after pill if they judge that they ought not do this because it would (in their view) facilitate the killing of a person? And does a religious bakery owner have the right to refuse to make a cake for a queer couple's wedding because the owner judges that her living a good or meaningful life involves her not being involved in queer weddings? If we grant that trans and nonbinary people's rights to change their gender markers on their identification documents are grounded in their rights to live and act with integrity, are we committed to holding that pro-life pharmacists and religious bakery owners have these rights?

I do not believe that we are so committed. One view of the boundaries of integrity-based rights, which I discussed briefly in section 1.1, is that the baker and the pharmacist have *pro tanto* integrity-based rights, but these *pro tanto* rights may be outweighed by the harms that their exercising these rights create.¹⁰⁷ On this view, my argument that trans and nonbinary people have *all-things-considered* integrity-based rights to change their gender markers that are not outweighed by any supposed harms does not imply that the pharmacist and baker have *all-things-considered* rights to refuse to sell the morning-after pill and to refuse to make a cake. For it might well be that there are bad consequences of the pharmacist and baker acting in line with their integrity—such as limiting people's access to abortions and the knock-on harms of this and the expressive harm done to queer people (see section 1.1). And it might well be that these bad consequences outweigh the pharmacist and baker's *pro tanto* integrity-based rights.

107 See, e.g., Billingham, "How Should Claims for Religious Exemptions Be Weighed?"

Furthermore, alternative views of rights to freedom of legal gender identification seem to be at least as likely to imply that the baker and pharmacist have these rights as the integrity-based account. If we hold that trans rights to change gender markers are autonomy based, then we similarly seem committed to the view that the pharmacist and baker have autonomy-based (*pro tanto*) rights to not be forced to sell the morning-after pill or to sell a wedding cake to a queer couple. And the pharmacist and baker might reasonably claim to be harmed by being forced to sell these things because they judge that they would be acting wrongly by selling these things, and this would harm them. Of course, we might hold that there is more harm done by allowing them to refuse services and therefore that a harm-based approach to rights to freedom of legal gender identification does not grant the pharmacist and the baker rights to refuse service. However, if we hold that granting them these rights would cause more harm than good, then we should hold that these bad consequences outweigh or defeat the baker's and pharmacist's *pro tanto* integrity-based rights such that they do not have all-things-considered integrity-based rights to refuse service. So a harm-based approach is not superior to an integrity-based approach in virtue of its implications regarding cases like these.

A different objection that I have encountered to my integrity-based approach is that autonomy, privacy, and freedom from harm might constitute integrity. However, living with integrity goes beyond living with autonomy and privacy and living free from harm. As discussed in section 1.1 above, to live with integrity we must live in line with

1. our practical identities;
2. our views of what gives our life meaning or what our life is "fundamentally about";
3. our views, commitments, or ideals regarding the kind of person we should be; and
4. the way of life that we value and take to be good for us, though not necessarily for everyone else.

Living in line with 1–4 goes beyond living with privacy and autonomy and living free from harm. Consider a case in which we need to do something *X* to live autonomously, to avoid harm, and to have privacy. Does this necessarily involve our needing to do *X* in order to live with integrity such that our living with integrity might be wholly constituted by our living with autonomy and privacy and avoiding harm? It does not seem so. Suppose that we strongly, deeply, and persistently desire not to have our age or height listed on our identification documents. The requirement that we list these things impinges on our autonomy, harms us, and plausibly impinges on our privacy too. But this does not

establish that we live without integrity if our age or height are listed on our identification documents, or that this requirement conflicts with our integrity. For we might just not want people knowing our age or height all the time or feel embarrassed if we are forced to present this information to others. This does not establish that we think that we cannot live a good or meaningful life while people know this information about us or while we are forced to present our age or height to others when we present our identification documents, or that we judge that we ought not present this information, or that our living in line with our practical identities requires that we not present this information to others. Our living in line with our integrity involves our living in line with our normative judgments. This involves something different from our living autonomously, with privacy, and free from harm. For we can be harmed and have our autonomy and privacy impinged upon without being forced to live out of line with our normative judgments (that is, out of line with 1–4)—for instance, without being forced to do something that we judge that we ought not do. So autonomy, privacy, and freedom from harm do not constitute integrity.

4. CONCLUSION

In this article, I first argued in section 1 that our basic liberal rights to live and act with integrity ground *pro tanto* rights to freedom of legal gender identification for trans and nonbinary people. I then argued in section 2 that trans and nonbinary people have all-things-considered rights to be able to change their gender markers on their legal identification documents relatively easily so that their gender markers do not clash with their gender identities. As I explained, there are several ways of realizing such all-things-considered rights, including self-identification policies, decertification policies, and policies (such as Scotland's recent proposal) that involve relatively modest wait periods on such changes of gender marker. So blanket bans on gender marker change—such as Hungary's, Kansas's, North Dakota's, Montana's, Oklahoma's, Florida's, Texas's, and the US federal government's—unjustly violate trans and nonbinary people's all-things-considered rights, as do policies—such as those in Arizona, Missouri, Nebraska, Wisconsin, and Singapore—that require trans people to demonstrate that they have had sex reassignment surgery before they change their gender markers. Finally, in section 3, I discussed alternative possible ways of thinking about trans and nonbinary people's rights to freedom of legal gender identification—namely, other possible grounds for these rights that can be gleaned from the literature on gender marker change in law and legal theory and the more general literature on trans rights. I showed that the integrity-based approach to the grounds of rights to freedom of legal

gender identification is superior to alternative approaches. Trans and nonbinary people have all-things-considered rights not to be forced to have gender markers on their legal identification documents that clash with their gender identities. These rights to freedom of gender are grounded in basic liberal rights. And many states currently act unjustly by breaching these trans rights to freedom of gender.¹⁰⁸

University of Leeds
r.cosker-rowland@leeds.ac.uk

REFERENCES

- Ables, Kelsey. "Florida Passes Bathroom Bill in Latest Wave of Anti-Trans Legislation." *Washington Post*, May 4, 2023.
- Ahmad, Nehaluddin, Ahmad Masum, and Abdul Mohaimin Ayus. "Freedom of Religion and Apostasy: The Malaysian Experience." *Human Rights Quarterly* 38, no. 3 (2016): 736–53.
- Andersson, Jasmine. "I Won't Even Be Allowed to Use My Name Now that Hungary Has Scrapped All Rights for Trans People." *The i Paper*, May 21, 2020. <https://inews.co.uk/news/i-wont-even-be-allowed-to-use-my-name-now-that-hungary-has-scrapped-all-rights-for-trans-people-429839>.
- Ashley, Florence. "Adolescent Medical Transition Is Ethical: An Analogy with Reproductive Health." *Kennedy Institute of Ethics Journal* 32, no. 2 (2022): 127–71.
- . "What Is It Like to Have a Gender Identity?" *Mind* 132, no. 7 (2023): 1053–73.
- . "'X' Why? Gender Markers and Non-Binary Transgender People." In *Trans Rights and Wrongs*, edited by Isabel C. Jaramillo and Laura Carlson. Springer, 2021.
- Audi, Robert. "Religious Liberty Conceived as a Human Right." In *Philosophical Foundations of Human Rights*, edited by Rowan Cruft, S. Matthew Liao, and Massimo Renzo. Oxford University Press, 2015.
- Australian Government. *Australian Government Guidelines on the Recognition of Sex and Gender*. July 2013, updated November 2015. <https://www.ag.gov.au/sites/default/files/2020-03/AustralianGovernmentGuidelinesontheRecognitionofSexandGender.pdf>.

¹⁰⁸ Thanks to audiences at the Australasian Association of Philosophy and Macquarie University. Thanks to Zoë Cosker, Stephanie Kapusta, and several reviewers for comments on previous versions of this article.

- Aziz, Muhammad Faisal Abdul. "Freedom of Religion by Religion for Religion." *Malaysia Now*, March 12, 2018. <https://www.malaysiakini.com/news/415372>.
- Barnes, Elizabeth. "Gender and Gender Terms." *Nous* 54, no. 3 (2020): 704–30.
- BBC News. "Fina Bars Transgender Swimmers from Women's Elite Events if They Went Through Male Puberty." June 20, 2022. <https://www.bbc.com/sport/swimming/61853450>.
- . "UK Athletics Wants Open Category for Male and Transgender Athletes." February 3, 2023. <https://www.bbc.com/sport/athletics/64514819>.
- Bettcher, Talia. "Evil Deceivers and Make-Believers: On Transphobic Violence and the Politics of Illusion." *Hypatia* 22, no. 3 (2007): 43–65.
- . "Through the Looking Glass: Trans Theory Meets Feminist Philosophy". In *The Routledge Companion to Feminist Philosophy*, edited by Ann Garry, Serene Khader, and Alison Stone. Routledge, 2017.
- Billingham, Paul. "How Should Claims for Religious Exemptions Be Weighed?" *Oxford Journal of Law and Religion* 6, no. 1 (2017): 1–23.
- Billson, Chantelle. "Cis Woman Harassed by Transphobe Who Followed Her into Female Toilet Because She Has Short Hair." *Pink News*, October 31, 2022. <https://www.thepinknews.com/2022/10/31/cis-woman-harassed-transphobe-female-toilet-short-hair/>.
- Bland, Archi. "Wednesday Briefing: The Scottish Trans Rights Law that Has Turned into a Constitutional Crisis." *Guardian*, January 18, 2023.
- Bou-Habib, Paul. "A Theory of Religious Accommodation." *Journal of Applied Philosophy* 23, no. 1 (2006): 109–26.
- Brighter, Cassie. "Trans-Later: Transitioning Late in Life." *Medium* (blog), May 19, 2020. <https://medium.com/empowered-trans-woman/trans-later-transitioning-late-in-life-d49d68b5213e>.
- Brooks, Libby. "'I've Been Spat On': Gender Non-Conforming Women Tell of Toilet Abuse After UK's Supreme Court Ruling." *Guardian*, August 13, 2025.
- Buss, Sarah, and Andrea Westlund. "Personal Autonomy." In *Stanford Encyclopedia of Philosophy* (Spring 2018). <https://plato.stanford.edu/entries/personal-autonomy/>.
- Cannoot, Pieter, and Ariël Decoster. "The Abolition of Sex/Gender Registration in the Age of Gender Self-Determination: An Interdisciplinary, Queer, Feminist and Human Rights Analysis." *International Journal of Gender, Sexuality and Law* 1, no. 1 (2020): 26–55.
- Chen, Heather. "Renouncing Islam in Malaysia Is Dangerous: We Spoke to Those Who Did It." *Vice*, April 2, 2021. <https://www.vice.com/en/article/pkd7gk/the-dangers-of-renouncing-islam-in-malaysia>.
- Chiam, Zhan, Sandra Duffy, Matilda González Gil, Lara Goodwin, and Nigel

- Timothy Mpemba Patel. "Trans Legal Mapping Report 2019: Recognition Before the Law." ILGA World, 2020. https://ilga.org/downloads/ILGA_World_Trans_Legal_Mapping_Report_2019_EN.pdf.
- Christman, John. "Autonomy in Moral and Political Philosophy." In *Stanford Encyclopedia of Philosophy* (Fall 2020). <https://plato.stanford.edu/entries/autonomy-moral/>.
- Clarke, Jessica. "They, Them, and Theirs." *Harvard Law Review* 132, no. 3 (2019): 894–991.
- CNN. "This Trans Influencer Received a Passport with the Wrong Gender After Trump's Executive Order." January 30, 2025. <https://www.cnn.com/2025/01/30/politics/video/trans-influencer-receives-wrong-gender-marker-on-her-passport-trump-executive-order-digvid>.
- Colburn, Ben. "Autonomy and Adaptive Preferences." *Utilitas* 23, no. 1 (2011): 52–71.
- Cook, Marceline. "10 Transgender People Share What They Wish They Knew Before Transitioning." *Self*, August 29, 2017. <https://www.self.com/story/before-transitioning>.
- Cooper, Davina, and Robyn Emerton. "Pulling the Thread of Decertification: What Challenges Are Raised by the Proposal to Refrom Legal Gender Status?" *Feminists@law* 10, no. 2 (2020): 1–36.
- Cooper, Davina, Robyn Emerton, Emily Grabham, Han Newman, Elizabeth Peel, Flora Renz, and Jessica Smith. "Abolishing Legal Sex Status: The Challenge and Consequences of Gender-Related Law Reform." King's College London, Future of Legal Gender Project Final Report, 2022. <https://www.kcl.ac.uk/law/research/future-of-legal-gender-abolishing-legal-sex-status-full-report.pdf>.
- Cosker-Rowland, Rach. "Gender Identity: The Subjective Fit Account." *Philosophical Studies* 181 (2024): 2701–36.
- . "Integrity and Rights to Gender-Affirming Healthcare." *Journal of Medical Ethics* 48, no. 11 (2022): 832–37.
- Cosker-Rowland, Rach, and Chris Howard. "Fittingness: A User's Guide." In *Fittingness*, edited by Rach Cosker-Rowland and Chris Howard. Oxford University Press, 2022.
- Davis, Heath Fogg. *Beyond Trans: Does Gender Matter?* NYU Press, 2017.
- Dembroff, Robin. "Real Talk on the Metaphysics of Gender." *Philosophical Topics* 46, no. 2 (2018): 21–50.
- Doran, Will. "Equality NC Director: No Public Safety Risks in Cities with Transgender Antidiscrimination Rules." *Politifact*, April 1, 2016. <https://www.politifact.com/factchecks/2016/apr/01/chris-sgro/equality-nc-director-no-public-safety-risks-cities/>.

- Dworkin, Ronald. *Justice for Hedgehogs*. Harvard University Press, 2011.
- . *A Matter of Principle*. Clarendon Press, 1985.
- . *Religion Without God*. Harvard University Press, 2013.
- . *Sovereign Virtue*. Harvard University Press, 2000.
- Elgot, Jessica. “Kemi Badenoch Could Rewrite Law to Allow Trans Exclusion from Single-Sex Spaces.” *Guardian*, April 5, 2023.
- Equality and Human Rights Commission (UK). “Separate and Single-Sex Service Providers: A Guide on the Equality Act Sex and Gender Reassignment Exceptions.” April 2022. <https://www.equalityhumanrights.com/sites/default/files/guidance-separate-and-single-sex-service-providers-equality-act-sex-and-gender-reassignment-exceptions.pdf>.
- Faye, Shon. *The Transgender Issue*. Penguin, 2021.
- Fletcher, Guy. *The Philosophy of Well-Being: An Introduction*. Routledge, 2016.
- Galop. “Trans Privacy Law.” June 10, 2021. <https://galop.org.uk/resource/trans-privacy-law/>. Archived August 28, 2023, at <https://archive.is/ZpssR>.
- Gerritse, Karl, Laura A. Hartman, Marijke A. Bremmer, Baudewijntje P. C. Kreukels, and Bert C. Molewijk. “Decision-Making Approaches in Transgender Healthcare: Conceptual Analysis and Ethical Implications.” *Medicine, Health Care and Philosophy* 24, no. 4 (2021): 687–99.
- Giordano, Simona. “Understanding the Emotion of Shame in Transgender Individuals: Some Insight from Kafka.” *Life Sciences, Society, and Policy* 14, no. 23 (2018): 1–22.
- Gogarty, Brendan. “All Colours of the Rainbow: Why Tasmania’s New Gender Identity Laws Are Warranted.” *Conversation*, June 21, 2020. <https://theconversation.com/all-colours-of-the-rainbow-why-tasmanias-new-gender-identity-laws-are-warranted-131664>.
- Grabham, Emily. “Decertifying Gender: The Challenge of Equal Pay.” *Feminist Legal Studies* 31, no. 2 (2023): 67–93.
- Guardian Sport. “Basketball Australia Rules Transgender Athlete Lexi Rodgers Ineligible to Play.” *Guardian*, April, 18, 2023.
- Halberstam, Jack. *Female Masculinity*. Duke University Press, 1998.
- Hanna, John. “Kansas Attorney General Sues to Prevent Transgender People from Changing Driver’s Licenses.” Associated Press, July 8, 2023.
- Hansler, Jennifer. “State Department Suspends Processing Passport Applications with ‘X’ Marker.” CNN, January 24, 2025.
- Hasenbush, Amira, Andrew R. Flores, and Jody L. Herman. “Gender Identity Nondiscrimination Laws in Public Accommodations: A Review of Evidence Regarding Safety and Privacy in Public Restrooms, Locker Rooms, and Changing Rooms.” *Sexuality Research and Social Policy* 16, no. 2 (2019): 70–83.

- Hayton, Debbie. "How the Trans Activists Fooled Ireland: Politicians Waved Legislation Through and the Public Isn't Happy." *UnHerd*, July 20, 2021. <https://unherd.com/2021/07/how-the-trans-activists-fooled-ireland/>.
- Hellman, Deborah. "Racial Profiling and the Meaning of Racial Categories." In *Contemporary Debates in Applied Ethics*, 2nd ed., edited by Andrew I. Cohen and Christopher Heath Wellman. Wiley-Blackwell, 2014.
- Hernandez, E. N., and Rowan Bell. "Much Ado About Nothing: Unmotivating 'Gender Identity.'" *Ergo* (forthcoming).
- Hill, Thomas. *Autonomy and Self-Respect*. Cambridge University Press, 1991.
- Hines, Sally. *Gender Diversity, Recognition and Citizenship: Towards a Politics of Difference*. Palgrave, 2013.
- HM Passport Office. "Guidance: Gender Recognition (Accessible)." April 10, 2024. <https://www.gov.uk/government/publications/gender-recognition/gender-recognition-accessible> (accessed August 21, 2025).
- ILGA Europe. "Annual Review of the Human Rights Situation of Lesbian, Gay, Bisexual, Trans and Intersex People in Europe and Central Asia: 2023." Brussels, February 2023. https://www.ilga-europe.org/sites/default/files/2023/full_annual_review.pdf.
- James, Sandy E., Jody L. Herman, Susan Rankin, Mara Keisling, Lisa Mottet, and Ma'ayan Anafi. "The Report of the 2015 US Transgender Survey." National Center for Transgender Equality, December 2016. <https://transequality.org/sites/default/files/docs/usts/USTS-Full-Report-Dec17.pdf>.
- Jenkins, Katharine. "Amelioration and Inclusion: Gender Identity and the Concept of Woman." *Ethics* 126, no. 2 (2016): 394–421.
- . *Ontology and Oppression*. Oxford University Press, 2023.
- Kapusta, Stephanie. "Gender Autonomy." In *The Routledge Handbook of Autonomy*, edited by Ben Colburn. Routledge, 2022.
- Kee, Caroline. "35 People Who Transitioned on How It Impacted Their Mental Health." *BuzzFeed News*, October 5, 2017. <https://www.buzzfeednews.com/article/carolinekee/people-talk-about-transitioning-and-mental-health>.
- Kelleher, Patrick. "Ireland Has Had Trans Self-ID Laws for Years." *Pink News*, February 1, 2023. <https://www.thepinknews.com/2023/02/01/trans-self-id-laws-gender-recognition-reform-scotland-ireland/>.
- Killmister, Suzy. *Taking the Measure of Autonomy*. Routledge, 2018.
- Laborde, Cecile. *Liberalism's Religion*. Harvard University Press, 2017.
- Lang, Nico. "Texas Just Quietly Revoked Trans People's Ability to Change Their Birth Certificates." *Them*, September 3, 2024. <https://www.them.us/story/texas-trans-birth-certificate-gender-marker-change-ban>.
- Lau, Holning. "Gender Recognition as a Human Right." In *The Cambridge Handbook of New Human Rights: Recognition, Novelty, Rhetoric*, edited by

- Andreas von Arnould, Kerstin von der Decken, and Mart Susi. Cambridge University Press, 2020.
- Lawford-Smith, Holly. "Ending Sex-Based Oppression: Transitional Pathways." *Philosophia* 49, no. 3 (2020): 1021–41.
- . *Gender Critical Feminism*. Oxford University Press, 2022.
- Lever, Annabelle. "Why Racial Profiling Is Hard to Justify: A Response to Risse and Zeckhauser." *Philosophy and Public Affairs* 33, no. 1 (2006): 94–110.
- Lopez, German. "Women Are Getting Harassed in Bathrooms Because of Anti-Transgender Hysteria." *Vox*, May 19, 2016. <https://www.vox.com/2016/5/18/11690234/women-bathrooms-harassment>.
- Maclure, Jocelyn, and Charles Taylor. *Secularism and Freedom of Conscience*. Harvard University Press, 2011.
- Middleton, Lucy. "Scotland's Trans Self-ID Bill No Risk for Women, Says UN Expert." Reuters, December 21, 2022. <https://www.reuters.com/article/britain-scotland-lgbt-idUSL8N33949W>.
- Ministry of Justice (UK), HM Prison and Probation Service, and the Right Honourable Dominic Rab. "New Transgender Prisoner Policy Comes into Force." Press release, February 27, 2023. <https://www.gov.uk/government/news/new-transgender-prisoner-policy-comes-into-force>.
- National Center for Trans Equality. "Summary of Birth Certificate Gender Change Laws." January 2020. <https://transequality.org/sites/default/files/docs/resources/Summary%20of%20State%20Birth%20Certificate%20Laws%20Jan%202020.pdf>.
- Nazri, Jeffry. "What Happens when Muslims in Malaysia Try to Leave Islam." *Medium* (blog), May 10, 2020. <https://jefrinazri.medium.com/what-happens-when-muslims-in-malaysia-try-to-leave-islam-339389970791>.
- Neuman Wipfler, A.J. "Identity Crisis: The Limitations of Expanding Government Recognition of Gender Identity and the Possibility of Genderless Identity Documents." *Harvard Journal of Law and Gender* 39, no. 2 (2016): 491–554.
- Nussbaum, Martha. *Liberty of Conscience: In Defense of America's Tradition of Religious Equality*. Basic Books, 2008.
- Office for Equality and Opportunity. "Response to Call for Input on Single-Sex Spaces: Guidance." December 17, 2024. <https://www.gov.uk/government/publications/response-to-call-for-input-on-single-sex-spaces-guidance/response-to-call-for-input-on-single-sex-spaces-guidance>.
- Oladipo, Gloria. "Majority of Trans Adults Are Happier After Transitioning, Survey Finds." *Guardian*, March 24, 2023.
- Pearce, Ruth, Sonja Erikainen, and Ben Vincent. "Introduction." In *TERF Wars: Feminism and the Fight for Transgender Futures*, edited by Ruth Pearl, Sonja

- Erikainen, and Ben Vincent. Sage Publishing, 2020.
- Plemons, Eric. *The Look of a Woman*. Duke University Press, 2017.
- Rajunov, Micah, and Scott Duane. *Nonbinary: Memoirs of Gender Identity*. Columbia University Press, 2019.
- Rawls, John. *Justice as Fairness: A Restatement*. Harvard University Press, 2001.
- Reed, Erin. "Anti-Trans Legislative Risk Assessment Map: July 2024 Edition." *Erin in the Morning* (blog), July 11, 2024. <https://www.erininthemorning.com/p/anti-trans-legislative-risk-assessment-3dc>.
- . "Kansas, Other States Threaten to Undo Legal Gender Changes: What To Do." *Erin in the Morning* (blog), June 28, 2023. <https://www.erininthemorning.com/p/kansas-other-states-threaten-to-undo>.
- . "Tennessee Law Rolls Back Trans Rights, Regressively Defines Sex." *Los Angeles Blade*, May 18, 2023. <https://www.losangelesblade.com/2023/05/18/tennessee-law-rolls-back-trans-rights-regressively-defines-sex/>.
- Renz, Flora. "Gender-Based Violence Without a Legal Gender: Imagining Singe-Sex Services in Conditions of Decertification." *Feminist Legal Studies* 31, no. 1 (2023): 43–66.
- . "Genders that Don't Matter: Nonbinary People and the Gender Recognition Act 2004." In *The Queer Outside in Law: Recognising LGBTQ People in the United Kingdom*, edited by Raj Senthurun and Peter Dunne. Palgrave, 2020.
- Renz, Flora, and Davina Cooper. "Reimagining Gender Through Equality Law: What Legal Thoughtways Do Religion and Disability Offer?" *Feminist Legal Studies* 30, no. 6 (2022): 129–55.
- Reuters. "England's Rugby Union and Rugby League Ban Transgender Players from Women's Game." July 30, 2022. <https://www.reuters.com/lifestyle/sports/rfu-bans-transgender-women-participating-womens-game-2022-07-29/>.
- Roan, Dan. "British Triathlon Becomes First UK Sport to Create 'Open' Category for Transgender Athletes." BBC News, July 5, 2022. <https://www.bbc.com/sport/triathlon/62063359>.
- Roche, Juno. *Trans Power*. Jessica Kingsley, 2020.
- Rummler, Orion. "Florida Is Quietly Denying Transgender Residents Updated Birth Certificates." *19th News*, July 15, 2024. <https://19thnews.org/2024/07/florida-transgender-updated-birth-certificates/>.
- Samuri, Mohd al Adib, and Muzammil Quraishi. "Negotiating Apostasy: Applying to 'Leave Islam' in Malaysia." *Islam and Christian-Muslim Relations* 25, no. 4 (2014): 507–23.
- Savulescu, Julian. "Ten Ethical Flaws in the Caster Semenya Decision on Inter-sex in Sport." *Conversation*, May 10, 2019. <https://theconversation.com/ten>

- ethical-flaws-in-the-caster-semenya-decision-on-intersex-in-sport-116448.
- Serano, Julia. "Transgender People, Bathrooms, and Sexual Predators: What the Data Say." *Medium* (blog), June 8, 2021. <https://juliaserano.medium.com/transgender-people-bathrooms-and-sexual-predators-what-the-data-say-2f31ae2a7c06>.
- . *Whipping Girl: A Transsexual Woman on Sexism and the Scapegoating of Femininity*. Seal Press, 2016.
- Spade, Dean. *Normal Lie: Administrative Violence, Critical Trans Politics, and the Limits of Law*. Duke University Press, 2015.
- Stavrou, Athena. "Trans Women to Be Banned from Single-Sex Spaces Under New EHRC Guidance." *Independent*, August 8, 2025.
- Steinmetz, Katy. "Why LGBT Advocates Say Bathroom 'Predators' Argument Is a Red Herring." *Time*, May 2, 2016. <https://time.com/4314896/transgender-bathroom-bill-male-predators-argument/>.
- Stewart, Ian. "2 Women Charged with Sexual Battery of Trans Woman in North Carolina Bar." *NPR*, January 9, 2019. <https://www.npr.org/2019/01/09/683711899/two-woman-charged-in-alleged-attack-on-trans-woman-in-north-carolina-bar>.
- Stock, Kathleen. *Material Girls: Why Reality Matters for Feminism*. Fleet, 2021.
- Stryker, Susan. *Transgender History*. Seal Press, 2008.
- Sunstein, Cass. "On the Tension Between Sex Equality and Religious Freedom." In *Toward a Humanist Justice: The Political Philosophy of Susan Moller Skin*, edited by Debra Satz and Rob Reich. Oxford University Press, 2009.
- Trevor Project. "National Survey of LGBTQ Youth Mental Health (2020)." New York, 2020. <https://www.thetrevorproject.org/wp-content/uploads/2020/07/The-Trevor-Project-National-Survey-Results-2020.pdf>.
- Theil, Michele. "Trans Americans Accuse Trump of 'Travel Ban': 'They Will Not Give Me Any Passport with Any Name.'" *Pink News*, January 31, 2025. <https://www.thepinknews.com/2025/01/31/trans-americans-passport-ban-trump-state-department/>.
- To, Margaret, Qi Zhang, Andrew Bradlyn, et al. "Visual Conformity with Affirmed Gender or 'Passing': Its Distribution and Association with Depression and Anxiety in a Cohort of Transgender People." *Journal of Sexual Medicine* 17, no. 10 (2020): 2084–92.
- Triggle, Nick. "Puberty Blockers for Under-18s Banned Indefinitely." *BBC News*, December 12, 2024. <https://www.bbc.com/news/articles/clyzzogx3p50>.
- Venditti, Valeria. "Gender Kaleidoscope: Diffracting Legal Approaches to Reform Gender Binary." *International Journal of Gender, Sexuality and Law* 1, no. 1 (2020): 56–75.
- Vincent, Ben. *Transgender Health*. Jessica Kingsley, 2018.

- Violet, Mia. "The Fact I Can't Marry as a Bride Is Another Reminder of How Unequal Trans Rights Still Are." *The i Paper*, February 24, 2020. <https://inews.co.uk/opinion/marry-bride-reminder-unequal-trans-rights-still-are-400587>.
- . *Yes, You Are Trans Enough: My Transition from Self-Loathing to Self-Love*. Jessica Kingsley, 2018.
- Warr, Megan. "How Do Gender Non-Conforming Individuals Experience Gendered Public Toilets?" BA diss., Coventry University, 2022. https://assets.ctfassets.net/x5yvfqv1tm38/3ETfh9BReNJ09hCjQHFSrZ/c2fo899ba40b33a9973b68b3745632b2/Megan_Warr_S__6057HUM_CW_submission.pdf.
- Weiss, Jillian Todd. "The Gender Caste System: Identity, Privacy, and Heteronormativity." *Law and Sexuality* 10 (2001): 123–86.
- Weiss, Suzannah. "9 Things People Get Wrong About Being Nonbinary." *Teen Vogue*, February 15, 2018. <https://www.teenvogue.com/story/9-things-people-get-wrong-about-being-non-binary>.
- Williams, Bernard. "A Critique of Utilitarianism." In *Utilitarianism: For and Against*, edited by Bernard Williams and J.J. C. Smart. Cambridge University Press, 1973.
- Williams, Rachel Anne. "What It Means to Be Authentic." *Medium* (blog), October 13, 2018. <https://medium.com/@transphilosophr/what-it-means-to-be-authentic-dbob8df40e50>.
- Yurcaba, Jo. "Florida Bars Transgender People from Changing the Sex on Their Driver's Licenses." NBC News, January 31, 2024. <https://www.nbcnews.com/nbc-out/out-news/florida-transgender-drivers-license-sex-change-gender-identity-rcna136395>.

WHY CONTRACTUALISM CANNOT ACCEPT EQUAL TREATMENT FOR EQUAL STATISTICAL LOSS

Jay Zameska

THE CONTRACTUALISM of T. M. Scanlon is a prominent nonconsequentialist approach to the ethics of risk, resulting in a significant and growing body of literature on contractualism and risk.¹ Contractualism holds that an action is morally permissible only if it is justified by principles that no one could reasonably reject. Unlike most consequentialist theories, however, contractualism rejects interpersonal aggregation in moral reasoning. Instead, it requires comparing the strength of individual personal reasons to determine which principles can be reasonably rejected. As such, a very strong reason of one person cannot be outweighed by combining many weaker reasons of other persons.² The goal is to identify “the principle whose implications are most acceptable to the person to whom it is least acceptable.”³

Debates over contractualism and risk typically make a distinction between *ex ante* and *ex post* variants of the theory. *Ex ante* contractualism evaluates principles based on individuals’ prospects under the proposed principle, which allows discounting the value of the expected outcome by its improbability. Here, however, my focus is specifically on *ex post* versions of contractualism, which do not allow discounting benefits and burdens by their improbability. Instead, they require considering the full value of the outcome, regardless of the probability of it occurring. In the rest of the article, unless otherwise specified, by ‘contractualism,’ I mean specifically the *ex post* variant of Scanlon’s contractualism.⁴

1 See Scanlon, *What We Owe to Each Other*. For discussion of contractualism and risk, see, *inter alia*, Hayenhjelm and Wolff, “The Moral Problem of Risk Impositions”; Fried, *Facing Up to Scarcity*; Ashford and Mulgan, “Contractualism”; and John and Curran, “Costa, Cancer and Coronavirus.”

2 Scanlon, *What We Owe to Each Other*.

3 Kumar, “Risking and Wronging,” 35.

4 For the distinction between *ex ante* and *ex post* contractualism, see Frick, “Contractualism and Social Risk”; and Suikkanen, “*Ex Ante* and *Ex Post* Contractualism,” which also

When addressing risky cases, some argue that *ex post* contractualism should comply with the following principle:

Equal Treatment for Equal Statistical Loss: We should treat cases alike if in both cases there is the same expectation of statistical loss and the only difference is the distribution of possible losses across possible outcomes.⁵

From here on, I refer to this as the *equal treatment principle*. Despite being an intuitively plausible principle, in this article, I argue that contractualism cannot comply with the equal treatment principle. This is because contractualism is a fundamentally *relational* moral theory.⁶ In brief, this means that contractualism is premised on the value of standing in a moral relationship with others. As such, contractualist moral deliberation is primarily not about assessing outcomes but about determining whether an action is justifiable to the individuals it affects. To structure this deliberation, contractualism represents those to whom we owe justification through *standpoints*. Standpoints are the generic positions from which individuals assess the burdens imposed by moral principles when determining whether those principles can be reasonably rejected. Importantly, standpoints are merely epistemic heuristics to help represent others and do not have normative importance themselves.⁷ This entails that unoccupied standpoints should not be considered in moral deliberation.

On this basis, I argue that we should draw a distinction between *definite standpoints*—those that we know with certainty pertain to at least one actual

develops a hybrid model. For defenses of *ex post* contractualism in the context of risk, see Reibetanz, “Contractualism and Aggregation,”; and R  ger, “On *Ex Ante* Contractualism.” Finally, for a broad overview of the *ex post* versus *ex ante* debate, see Ashford and Mulgan, “Contractualism.” While the choice between these frameworks is a major debate, often adjudicated by cases like the ones discussed here, the aim of this article is more modest. It seeks to explore a question that is internal to the *ex post* framework: whether the theory’s own foundational commitment to relational justification has unexamined consequences for how it must treat principles governing risk distribution. Understanding these consequences on the theory’s own terms, I suggest, is a necessary step that is logically prior to the larger debate.

- 5 Steuwer, “Contractualism, Complaints, and Risk,” 119. I take Steuwer’s wording and discussion of this as my focus because they are a particularly clear and explicit example of this principle. But this general idea lies behind a significant amount of discussion, especially in opposition to the identified victim bias. I return to discussion of the identified victim bias in section 3.
- 6 Scanlon, *What We Owe to Each Other*.
- 7 See, *inter alia*, Scanlon, *What We Owe to Each Other*, ch. 5; Gibb, “Relational Contractualism and Future Persons”; Katz, “Contractualism, Person-Affecting Wrongness and the Non-Identity Problem”; and Martin, “Navigating Nonidentity.”

person—and *indefinite standpoints*—those about which we are uncertain whether they pertain to any actual individuals. I argue that given contractualism's relational foundation, we should respond differently to definite and indefinite standpoints. We should, in general, prioritize the reasons of definite standpoints over those of indefinite standpoints. This in turn demonstrates that the equal treatment principle cannot always be correct: cases that are probabilistically equal may in fact differ in their justifiability and are thus morally different, from a contractualist perspective.

To develop this argument, in section 1, I introduce the distinction between definite and indefinite standpoints in more detail, and explain why it follows from contractualism's relational foundation. In section 2, I present an example case to demonstrate that even when two choices involve equal expectations of statistical loss, their justifiability may differ depending on whether they involve definite or indefinite standpoints. I argue that, in the case at hand, prioritizing definite standpoints is necessary to avoid acting unjustifiably, and thus, the principle of equal treatment cannot be right. This shows that cases with equal expected statistical losses cannot always be treated equally. In section 3, I address several objections, including concerns about the distinction between objective and epistemic risk, the potential implications of rejecting the equal treatment principle, challenges to the requirement that contractualism requires justifiability actions to actually existing individuals, as opposed to merely hypothetical ones, and the distinction between act and rule contractualism. Finally, in section 4, I consider the broader implications, including how this argument provides a contractualist justification for the identified victim bias and a possible framework for *ex post* discounting.

1. STANDPOINTS, UNCERTAINTY, AND INDIVIDUAL JUSTIFICATION

Consider two possibilities:

Gun: There is a person and a revolver in front of me. The revolver's six-chamber cylinder has already been loaded with a single round, and the cylinder has already been spun. I can pick up the revolver, point it at the person, and pull the trigger one time.

Box: There is an opaque box in front of me. There is a one-sixth chance that someone has been placed inside the box. I can press a button and destroy the box (and anyone inside).

Assuming that I must pick one of these two possibilities, which should I choose? At the moment, it may not seem to matter which option I choose. They

seem to be equal in all relevant respects.⁸ From an *ex ante* perspective, there is a one-sixth chance of killing someone, regardless of which option I choose. From an *ex post* perspective, there is one single death to consider, regardless of which option I choose. As such, Gun versus Box seems to be precisely the kind of dilemma to which the equal treatment principle should apply. In both cases, the expected statistical loss is equal: in one out of six possible worlds, a single person dies regardless of whether I choose Gun or Box. The only thing that changes is the distribution of deaths across designators (the victim of Gun versus the person in the opaque box).

Given this equal expected statistical loss, the equal treatment principle holds that we should treat the two options the same. Perhaps we should flip a coin, perhaps we should simply think either choice is permissible, or perhaps we should employ some sort of tiebreaker.⁹ Here, I assume we should flip a coin, although ultimately it does not matter to this argument how contractualists should respond to equal cases: Gun and Box are probabilistically equivalent, but they are not morally equal for contractualists, and thus, they should not be treated equally.

This is because contractualism is fundamentally a *relational*, or *second-personal*, moral theory.¹⁰ As Rahul Kumar explains, “the contractualist claim is that all persons stand in a particular kind of relationship to one another, the

8 One might worry that from an *ex ante* perspective, the cases as described may not be equal. Gun appears to impose a concentrated risk on a specific person, while Box could be seen as imposing a trivial risk dispersed over a large population. While this article's focus is on the *ex post* complaints, it is possible to stipulate a background structure that may neutralize this concern. We can imagine that the potential victims for both scenarios are drawn from two separate, equally large populations, such that the initial *ex ante* chance of being the one ultimately harmed is identical in both setups. For Gun, one person is selected from a population and placed in the role of ‘victim of Gun’. For Box, a different person is selected from the other population to be the potential ‘person in the box’, who is then subject to the one-sixth chance of actually being placed inside. Admittedly, from an *ex ante* perspective that employs Frick's decomposition test, this equalizing setup may not be convincing, as such a view focuses on the unequal risk distributions at the immediate moment of choice, regardless of previous risk distributions. From the strictly *ex post* perspective of this article, however, what matters is that both setups entail an equal expected statistical loss. As such, the moral asymmetry at the moment of decision—the choice between a definite and an indefinite standpoint—remains, and it is this *ex post* feature that my argument addresses. My thanks to an anonymous reviewer for pushing me to clarify this point.

9 See Taurek, “Should the Numbers Count?”; Wasserman, “Let Them Eat Chances”; Saunders, “A Defence of Weighted Lotteries in Life Saving Cases”; Scanlon, *What We Owe to Each Other*; and Reibetanz, “Contractualism and Aggregation.”

10 Scanlon, *What We Owe to Each Other*. For more general discussion of second-personal moral theory, see also Darwall, *The Second-Person Standpoint*.

ideal form of which realizes a distinct value, that of mutual recognition.”¹¹ As such, contractualism is premised on the value of standing in a moral relationship to others, and it is the value of this relationship that requires us to act in ways that others could not reasonably reject.¹² Not only does this relational foundation help define some of the theory’s most controversial and distinctive commitments (for example, its famous ban on aggregation), but I argue that it also means that we must treat Gun and Box differently—and thus, the equal treatment principle cannot always be right. Although Gun and Box may be probabilistically equal, they represent a more fundamental difference in contractualist theory regarding how contractualists should address two different kinds of standpoints, which I call *definite* and *indefinite* standpoints. To motivate the relevance of this distinction, first I explain the concept of standpoints in contractualism in more depth, before explaining why contractualism must treat them differently.

Standpoints are the generic positions from which individuals assess the burdens imposed by a given principle when determining whether that principle can be reasonably rejected. An important but often overlooked feature of such standpoints is that they are intended to serve only as a kind of *epistemic heuristic* that allows us to assess the claims of actually existing but currently unidentified others.¹³ As Corey Katz explains, “the value at the heart of contractualist metaethics is the value of living in *actual second-personal relationships* of mutual justifiability,” and as a result, “what matters for an agent is that her action be justifiable *to the actual people she is in relationship with* on reasons those people could not reasonably reject.”¹⁴ It is the value of these moral relationships, constituted by a shared capacity for rational agency, rather than the reasons available to standpoints per se, that underwrites the contractualist focus on justifying our actions to others on grounds they could not reasonably reject.¹⁵

11 Kumar, “Risking and Wronging,” 258.

12 Scanlon, *What We Owe to Each Other*; and Southwood, “Moral Contractualism” and *Contractualism and the Foundations of Morality*.

13 See Gibb, “Relational Contractualism and Future Persons”; Katz, “Contractualism, Person-Affecting Wrongness and the Non-Identity Problem”; and Martin, “Navigating Non-identity”—all of which provide significant discussion of this epistemic heuristic aspect of standpoints.

14 Katz, “Contractualism, Person-Affecting Wrongness and the Non-Identity Problem,” 112 (emphasis added).

15 It also is this focus on the moral relationships between rational agents rather than on mutual advantage or pure self-interest that leads Scanlon’s contractualism to often be described as a Kantian contract theory (as opposed to, e.g., Hobbesian contract theories). For discussion, see Southwood, “Moral Contractualism.”

In other words, standpoints are simply a way to overcome the limited information regarding the particular identities of the actual individuals our actions will affect.¹⁶ Given that they are limited to an epistemic role, standpoints have normative relevance only insofar as they represent *actual* individuals. This means that the moral relevance of the generic reasons associated with a standpoint is contingent on that standpoint being occupied by at least one actual person. Thus, when we are certain no one actually occupies a given standpoint, we can disregard that standpoint and its generic reasons, as “the reasons of a merely imagined but unoccupied standpoint have no bearing on any of our moral relationships, and it is the value of these relationships, and only these relationships, that drive our moral judgments.”¹⁷

In short, contractualism is fundamentally premised on the moral relationship between actual individuals. Standpoints are an epistemic heuristic to allow us to consider the reasons of actual but unidentified others.¹⁸ As such, they have no independent moral significance beyond the individuals they represent. This entails that when we are certain a standpoint is empty—that it does not correspond to an actual individual—we should not consider reasons from that standpoint. Conversely, if we know a standpoint is occupied—that is, if we know it corresponds to at least one actual individual—then we must consider the reasons from that standpoint.

I take both of the preceding points to be fairly uncontroversial and, at this stage, perhaps rather uninteresting. But in between “certainly empty” and “certainly occupied” are a range of standpoints that are morally interesting. These are standpoints where we do not know if they are in fact occupied. I call these *indefinite standpoints*, in contrast to *definite standpoints*, where we are certain that the standpoint is occupied. Definite standpoints have clear normative force because they represent actual persons to whom justification is owed. Indefinite standpoints, by contrast, represent a special case of uncertainty—where it is unclear whether anyone occupies the standpoint and therefore whether any moral relationship is affected by our decision. The uncertainty involved in each of these is fundamentally different. In a case like Gun, the uncertainty is about the outcome for a person we know exists; the standpoint is definite, but the result is not. In a case like Box, however, the uncertainty is more fundamental: we are uncertain about whether there is an actual person to whom justification could be owed in the first place. This introduces a distinct risk, unique to the

16 For the original discussion of this point, see Scanlon, *What We Owe to Each Other*, 202–6.

17 Gibb, “Relational Contractualism and Future Persons,” 12.

18 *Inter alia*, Scanlon, *What We Owe to Each Other*; and Gibb, “Relational Contractualism and Future Persons.” See also Martin, “Navigating Nonidentity.”

relational commitments of contractualism: the risk of acting on the basis of a merely possible standpoint to the detriment of a certainly existing one. As I argue in the next section, this would be to act in a way that is fundamentally and uniquely unjustifiable.¹⁹ This distinction between definite and indefinite standpoints matters because, as just explained, the normative relevance of a standpoint depends on whether it represents actual individuals. Contractualism's relational foundation requires justifiability to actual persons, and as such, indefinite standpoints introduce uncertainty about whether justification is possible at all. As I argue in the next section, this means that in the cases introduced earlier, the uncertainty in Box is different from the uncertainty in Gun, even if they are probabilistically equal and have equal expected statistical losses.

2. WHY EQUAL STATISTICAL LOSS DOES NOT GUARANTEE EQUAL JUSTIFIABILITY

With this discussion of standpoints in mind, let us return to the choice from earlier in this article.

Gun: There is a person and a revolver in front of me. The revolver's six-chamber cylinder has already been loaded with a single round, and the cylinder has already been spun. I can pick up the revolver, point it at the person, and pull the trigger one time.

Box: There is an opaque box in front of me. There is a one-sixth chance that someone has been placed inside the box. I can press a button and destroy the box (and anyone inside).

The idea that we should treat Gun and Box equally stems at least in part from the idea that both potential victims can present equally strong reasons against shooting the gun or destroying the box. In both cases, they can each offer the reason "I will die."²⁰ However, these reasons should not in fact be treated equally.

19 To make this distinction clearer, we can label these as two different kinds of risk. The uncertainty in Gun can be called *outcome risk*—uncertainty about the consequences of an action. The more fundamental uncertainty in Box is what I call *justificatory risk*—the risk that our action will be foundationally unjustifiable because a necessary condition for justification is absent. As I argue in the next section, contractualism should aim to avoid justificatory risk.

20 Note again, this is because I am focusing only on *ex post* contractualism, which does not allow discounting based on probability. Consequently, each person's complaint should represent the full disvalue of the possible negative outcome, regardless of its likelihood—in this case, death. Contrast this with the *ex ante* version of the complaint—namely, "I face a one-sixth chance of death."

To see why, it is helpful to consider the possible states of the world regarding Box:

S_1 : The box is full.

S_{2-6} : The box is empty.

Imagine that we are in S_1 . If the box is full, either choice (Gun or Box) would be justifiable. This is because presumably, if you can spare only one of two people from an equal harm but not both, and everything else is equal, it is justifiable to spare either one, as the reasons on either side are equally strong. It is a tragic choice, but it is not an unjustifiable one, and justification is what matters to contractualism. If we flip a coin, regardless of whether it is heads or tails, our resulting choice will be justifiable to both parties, and neither will be able to reasonably reject our decision.

Consider how things would change if we are instead in any of S_{2-6} . If we decide to destroy the box, our action would clearly be justifiable, for exactly the same reasons as in S_1 . But now imagine that we instead decide to shoot the gun. For this to be justifiable would introduce a dilemma: given that there is no one in the box, to justify our choice on the basis of the reason "I will die" from the person-in-the-box standpoint would require either that either (1) it is justifiable for reasons that have nothing to do with our moral relationship, as there is no one in the box with whom we have such a relationship or (2) we do in fact have a moral relationship with the empty person-in-the-box standpoint. This raises a dilemma for contractualism, either horn of which is unacceptable.²¹ The first horn is incompatible with the basic structure of the theory, and the second horn stretches the notion of the "moral relationship" to an implausible degree.

This dilemma arises because if the box is empty, there can be no individual, personal reasons of the kind that contractualism demands. The justification for choosing Gun in S_{2-6} (or choosing to flip a coin) would fail to be based on the right kind of reason, as there is no actual individual to hold such a reason. Alternatively, if we were to claim that there are such reasons, this entails that, in Gibb's terms, "we would have to claim that we share a moral relationship with something like a metaphysical entity known as a 'standpoint.'"²² In short, if the box is empty, the contractualist mode of justification breaks down. We cannot properly consider the reasons for rejection from the person-in-the-box standpoint because if the box is empty, they would not be the right kind of reasons

21 For the original discussion of this dilemma, see Gibb, "Relational Contractualism and Future Persons."

22 Gibb, "Relational Contractualism and Future Persons," 11.

required by contractualist moral deliberation, and as a result, we would act in a way that the victim in Gun could reasonably reject.

Thus, if we want to avoid the risk of acting unjustifiably, we must choose Box. Choosing Box does not necessarily minimize the risk of harm, but it minimizes the risk of acting unjustifiably. This is because there is a meaningful distinction between the risk of (justifiably creating) a bad outcome and the risk of acting unjustifiably in the first place; for contractualism, the latter is a more serious failure.²³ Given that contractualism requires that you act justifiably, it also requires that you act in ways that are, *ceteris paribus*, guaranteed to be justifiable over acting in ways that entail a risk of being unjustifiable. This means that the equal treatment principle cannot be right. In this case, we have equal statistical loss, but due to differences in standpoints, we cannot treat them equally. More than just highlighting a problem for the equal treatment principle, this argument also generalizes to any competitive choice between definite and indefinite standpoints with equal harms (or benefits). In competitive cases where both definite and indefinite standpoints face an equal burden, contractualists should not treat them equally. To do so risks acting unjustifiably, in an analogous way to choosing Gun over Box.

As such, in any case where we either prioritize the reasons of indefinite standpoints or treat them equally to definite standpoints (by, e.g., flipping a coin or engaging in some similar procedure), we run the risk of acting unjustifiably. In contrast, if we prioritize the reasons of definite standpoints, we never run the risk of acting unjustifiably. In other words, reasons from definite standpoints carry greater weight than those from indefinite standpoints. Here, I have limited my discussion to cases with equal harms in order to mirror the structure of the case above (and to address the equal treatment principle specifically). Given that in the case at hand, the only difference between the reasons under consideration is whether they stem from definite or indefinite standpoints, this greater weight is a decisive consideration. However, in the next section, I briefly return to the question of how to deal with this priority to definite standpoints when harms and benefits are not equal.

To summarize, in cases involving competitive choice between definite and indefinite standpoints, as in Gun versus Box, we must prioritize the interests of definite standpoints. Prioritizing the reasons of definite standpoints is the only way to avoid the risk of acting unjustifiably. If the indefinite standpoint turns out to be unoccupied, it cannot provide reasons of the kind demanded by

23 This maps onto the distinction drawn earlier between what I call outcome risk and justificatory risk. The risk of a bad outcome (harm) is an outcome risk, present in both Gun and Box. The risk of acting unjustifiably, however, is justificatory risk, introduced by the indefinite standpoint in Box.

contractualism as a fundamentally relational moral theory. As such, acting on the basis of such reasons against the interests of another, actual person is unjustifiable. Given that contractualism requires us to act justifiably, it also requires us to choose actions that are certainly justifiable over actions that are “riskily” justifiable, all else equal. Consequently, in competitive cases with equal burdens, contractualists must always prioritize the interests of definite standpoints over indefinite standpoints, and thus, the equal treatment principle should be rejected. Next, I consider four major objections before turning to the broader implications of this argument.

3. OBJECTIONS: OBJECTIVE VERSUS EPISTEMIC RISK, IMPLAUSIBLE CONSEQUENCES, THE ROLE OF HYPOTHETICAL JUSTIFIABILITY, AND ACT VERSUS RULE CONTRACTUALISM

In this section I address four objections—namely, (1) that this argument confuses or otherwise neglects the distinction between objective and epistemic risk, risk concentration/dispersal, or the distinction between the *luckless* and the *doomed*; (2) that this argument against the principle of equal treatment cannot be right because it entails implausible consequences; (3) that the requirement that contractualism requires justifiability to actually existing people is incorrect; and (4) that the argument is limited to “act contractualism” and fails when applied to the traditional rule-based formulation of the theory. I address each of these in turn.

First, some may object that I am drawing a distinction that already has a more common name: *objective risk* versus *epistemic risk*. In other words, the objection holds that there is already an established difference between Gun and Box, and it has nothing to do with standpoints but rather has to do with whether the risk is objective or merely epistemic. In one common way to cast this distinction, objective risk refers to “truly” risky cases, whereas epistemic risk refers to cases that are risky simply because of our lack of knowledge.²⁴ Some have sought to support the idea that these different kinds of risk matter morally and thus that perhaps we ought to respond to them differently, even if they have the same probabilities.²⁵

24 For general discussion of this objective-versus-epistemic split in interpretations of probability, see Gillies, *Philosophical Theories of Probability*.

25 For discussion see, *inter alia*, Tadros, “Controlling Risk”; Oberdiek, *Imposing Risk*; Fried, *Facing Up to Scarcity*; and Spiekermann, “Good Reasons for Losers.” For discussion of this distinction specifically in the context of contractualist approaches to risk, see Otsuka, “Risking Life and Limb”; Fried, *Facing Up to Scarcity*; and Steuwer, “Contractualism, Complaints, and Risk.”

Along these lines, some may object that Gun is objectively risky while Box is merely epistemically risky. The objection would hold that at the time of decision, it is already decided whether there is someone in the box, but we simply do not know. In contrast, spinning the cylinder in the revolver in Gun is “truly” risky. As such, it may be that these cases are not in fact equal. This is not quite right, however, as assuming there is some mechanism similar to spinning the revolver’s cylinder to decide whether someone is in the box, in both Gun and Box, we are dealing with objectively risky situations—at least insofar as there are such things as objective risks.

Some may still insist that the objective risk in Box occurs *before* you make your decision, and so it truly is a case of objective risk (Gun) versus epistemic risk (Box). At the time of decision, it is either the case that there is someone in the box, or it is the case that there is not—so it cannot be objectively risky in the way that Gun is. Instead, it is merely epistemically risky because it is based on a lack of knowledge of the contents of the box. However, this way of casting the objection does not succeed either. Gun is at the time of trigger pull *also* purely epistemic. It has already been decided which chamber the bullet is in, I simply do not know which. So either both Gun and Box are epistemic, or both are objective, depending on the precise slice of time we select. As such, the difference between epistemic and objective risks cannot be what makes a moral difference here. Instead, as discussed earlier, what matters is whether the standpoint is definite or indefinite.

In other words, either both Gun and Box are objectively risky, or both are epistemically risky. At the time of decision making, it is already decided whether there is a bullet in the firing chamber, just as it is already decided whether there is someone in the box. So the case is merely epistemically risky either way. Or we could partition things differently and instead insist that it is objectively risky, as in the preceding stage of each case, there is an objective one-sixth chance of the bullet ending up in the firing chamber and an objective one-sixth chance of placing a person in the box. The underlying point is that we can shift whether it is objectively or epistemically risky based on how we partition time, but in either case, it applies equally to both Gun and Box. What makes the difference is whether our actions are justifiable to everyone involved at the time of decision, and as argued above, this turns in part on whether we are considering the reasons of definite or indefinite standpoints.

To put the point more directly, the definite/indefinite distinction cuts across the epistemic/objective one by focusing on a different feature of the situation. The moral difference stems not from the nature of the risk (epistemic/objective) but from what it is we are uncertain about. In both cases, we are uncertain whether our action will result in someone’s death; this is a risk

concerning the outcome for a person. But exclusively in Box, we are also uncertain whether our action can be justified at all because we are uncertain whether the person-in-the-box standpoint is actually occupied.²⁶ This second form of uncertainty is unique to a relational moral theory like contractualism, which is grounded in the idea of justifiability to actual, existing individuals. It is this latter uncertainty about the grounds of justification itself, not the epistemic nature of the risk, that makes the cases morally different.

A related objection holds that the definite/indefinite distinction is simply a proxy for the more familiar distinction of risk concentration versus dispersal. One might argue that a preference for Box over Gun is easily explained by the fact that the former seems to distribute risk broadly, while the latter concentrates it on a single individual, thus making the definite/indefinite distinction redundant.²⁷ However, to show that the definite/indefinite distinction is conceptually separate from concerns about risk distribution, it is useful to consider a hypothetical case where the two principles come into direct conflict. Imagine a choice between (1) imposing a large, concentrated risk on a single indefinite standpoint (e.g., a high probability of harm to a person whose existence is uncertain) and (2) imposing a small, dispersed risk across many definite standpoints (e.g., a low probability of harm to many actual, identified people). A principle focused only on avoiding concentrated risk may favor option 2. The argument of this article, however, provides a strong reason to prefer option 1. This is because choosing 2 involves burdening actual people for the sake of a merely possible one, risking a failure of justifiability, whereas choosing 1 avoids this particular kind of relational failure. That the two principles can yield conflicting recommendations proves that they are conceptually distinct.

A final version of this objection holds that the distinction I draw between definite and indefinite standpoints is reducible to another existing distinction: rather than a relabeling of objective versus subjective, it may also appear to be a relabeling of “doomed” and “luckless” victims.²⁸ On this view, a doomed victim is one whose fate is already determined, even if it is epistemically unknown,

26 This corresponds to the two kinds of risk discussed earlier. The uncertainty about death can be described as outcome risk. The second, more fundamental uncertainty about whether the standpoint is occupied—and thus whether our action is justifiable—is what I call justificatory risk. It is the presence of this second kind of risk in Box that makes the case morally different for contractualism.

27 My thanks to an anonymous referee for suggesting this objection. It is worth noting, however, that the moral significance of risk concentration is itself a matter of debate. For skepticism on this point, see Eyal, “Concentrated Risk, the Coventry Blitz, Chamberlain’s Cancer.”

28 I thank an anonymous referee for raising this objection and pushing me to clarify the relationship between my account and this distinction. For discussion of this distinction,

while a luckless victim faces a genuinely probabilistic risk. The objection is that my argument for prioritizing definite standpoints is simply another way of arguing that we should prioritize doomed victims.²⁹

While the two may often align, they are grounded in different moral considerations. The doomed/luckless distinction is about the causal nature of the risk that an individual faces. My account, in contrast, is grounded in the relational requirements of contractualism; the definite/indefinite distinction turns on our certainty about a standpoint's occupation and the corresponding risk of justificatory failure. To see how these justifications come apart, consider the following variation on Gun.

Definite but Luckless: A revolver is pointed at a person, but the trigger is now connected to a truly indeterministic quantum randomizer with a one-sixth chance of firing one time.

The victim in this scenario is luckless, not doomed. A theory based on prioritizing the doomed would not grant this person special priority. Yet on my account, the victim's standpoint is still definite—we are certain an actual person occupies that role—and thus their reasons take priority over the indefinite standpoint in Box to avoid justificatory risk.

Furthermore, my distinction is better equipped to explain the moral difference in the original Gun versus Box decision. Arguably, both victims could be classified as doomed, since the chancy portions of the relevant causal chains (the spin of the revolver's cylinder, the sorting into the box) have already been completed. The doomed/luckless distinction therefore struggles to differentiate them. The definite/indefinite distinction, however, clearly separates the cases based on the certainty of standpoint occupation. Thus, the justification for prioritizing definite standpoints is a distinct, relational one, not based on the nature of the risk that an individual faces.

Perhaps the strongest objection is that this argument may seem to force contractualism into a position where it must endorse implausible conclusions. For example, consider a new choice that must be made.

Gun: There is a person and a revolver in front of me. The revolver's six-chamber cylinder has already been loaded with a single round, and the cylinder has already been spun. I can pick up the revolver, point it at the person, and pull the trigger one time.

see Otsuka, "Risking Life and Limb"; and Steuwer, "Contractualism, Complaints, and Risk."

29 For discussion of prioritizing between doomed and luckless victims, see Steuwer, "Contractualism, Complaints, and Risk."

Box (Massive): There is an opaque box in front of me. There is a one-sixth chance that one hundred people have been placed inside the box. I can press a button and destroy the box (and anyone inside).

The view I defend seems to hold that contractualists should choose to potentially sacrifice all one hundred people in *Box (Massive)* rather than risk a single fatal shot in *Gun*. After all, the first is a definite standpoint, whereas each of the one hundred others is indefinite, and I argue that contractualism should prioritize definite over indefinite standpoints. In other words, my argument seems to require potentially sacrificing one hundred lives to spare a single individual a one-sixth risk of death.³⁰ And this seems extremely implausible.

Notably, this objection does not defend the equal treatment principle or undercut the argument against it. It just shows contractualism to sometimes yield putatively implausible conclusions. However, the equal treatment principle does not prevent this kind of implausible conclusion. In fact, the equal treatment principle does not apply because the expected statistical loss is not equal. As such, this objection does not undermine this argument against the equal treatment principle. Instead, this objection represents a problem inherent to contractualism's anti-aggregative commitments rather than a specific issue with how it handles expected statistical losses. However, this is a more general problem for contractualism and is not new.³¹ There are responses in the contractualist literature, ranging from Scanlon's tiebreaking argument to weighted lotteries.³² Since this is beyond the scope of this article, I do not address this substantial body of literature here.

As such, the argument presented here is not shown wrong by this objection, as this implausible conclusion is a consequence of the structure of contractualism, and similar implausible conclusions can be constructed without reference to risk, uncertainty, or expected statistical loss.³³ Consequently, my argument against the principle of equal treatment may in fact also work as an argument

30 This phrasing sets aside the fact that I am looking at this in terms of *ex post* contractualism, which requires viewing this as one death versus one hundred deaths rather than as a one-sixth chance of one death versus one hundred deaths. But the objection is sharper and stronger when considering it in *ex ante* terms.

31 For an overview and critical discussion of contractualism's difficulties with aggregation, see Fried, *Facing Up to Scarcity*.

32 Scanlon, *What We Owe to Each Other*; and Saunders, "A Defence of Weighted Lotteries in Life Saving Cases."

33 See, for example, discussions over whether contractualists should save a single person or some larger number of people from death: *inter alia*, Taurek, "Should the Numbers Count?"; Otsuka, "Scanlon and the Claims of the Many Versus the One"; Fried, *Facing Up to Scarcity*; and Muñoz, "Each Counts for One."

against contractualism itself, if opposition to this conclusion is sufficiently strong. The fact that contractualism cannot endorse the equal treatment principle and will at least sometimes license sacrificing one hundred lives to avoid a one-sixth chance of death for a single individual may be recast as an objection to contractualism itself.³⁴ Thus, this implausibility objection does not actually challenge the argument against the equal treatment principle, although it may change its strategic use.

Finally, my argument is premised on a strict interpretation of contractualism's relational foundation, which holds that justification is owed to *actual* individuals even if the process of justification itself is idealized and hypothetical. This view, however, is not the only possible understanding, as there is ongoing debate in contractualism over exactly how idealized the process of justification should be. A contrasting, more strongly idealized view holds that the process of justification can be detached from actual persons. On this view, what matters for contractualist moral deliberation are the reasons associated with a standpoint itself, not whether that standpoint is currently occupied.

Against this highly idealized position, my argument defends the stricter, actualist interpretation. While contractualism is about justifiability rather than actual justification, I contend that what matters is justifiability to *actual individuals* rather than justifiability in a purely hypothetical sense.³⁵ As such, we need not justify ourselves to people who we know do not and will not exist. To see this, consider a case where there is equal expected loss between a standpoint we know is certainly occupied and one we know is certainly empty. It would be wrong to treat them equally on the grounds that contractualism is concerned with justifiability rather than justification. Even though the concern is justifiability, it is justifiability to *actual people* that matters. To deny this would be to reject the fundamentally relational basis of the theory.

Furthermore, contractualism's oft discussed *impersonalist restriction* gives another reason to doubt the idea that contractualism is concerned with justifiability to merely hypothetical people. The impersonalist restriction holds that "impersonal values are not themselves grounds for reasonable rejection," and as such, only personal reasons may be employed for contractualist justification.³⁶

34 It may alternatively be read as an objection specifically to the various features of contractualism that permit this rather than to contractualism as a whole. Some have already advocated for removing or revising the individualist restriction on similar grounds (e.g., Parfit, *On What Matters*), but the argument here could also be used to argue for revising the relational character of contractualism.

35 Although for an argument that contractualism should focus on actual justification, not justifiability, see Kim, "On the Need for Real Dialogue."

36 Scanlon, *What We Owe to Each Other*, 222.

However, the only kind of reasons that merely hypothetical and nonexistent individuals could have for reasonable rejection are impersonal reasons. They cannot be personal reasons because they are not the reasons of any (actual) individual. As such, if they are reasons at all, they must be impersonal reasons and are thus inadmissible in a contractualist framework. So the basic commitments of the theory speak against the idea that contractualism should consider merely hypothetical individuals. Nonexistent people cannot provide reasons of the right type.³⁷

The fourth and final objection I consider draws on recent discussions of the distinction between *act contractualism* and the more traditional *principle contractualism*, or *rule contractualism*.³⁸ In this framework, principle contractualism holds that an act is wrong if it violates a principle that no one could reasonably reject for the general regulation of behavior, whereas act contractualism assesses the wrongness of an act based on its own rejectability, independent of general principles.³⁹ The objection holds that the argument presented earlier may apply to the act-based version of the theory but not to the two-stage, rule-based contractualism initially put forward by Scanlon. In the context of the Gun versus Box cases introduced earlier, under the standard rule-based version of contractualism, we would assess not the single choice between Gun and Box but rather competing general principles or rules for governing such situations. For example, we might choose between a principle of equal treatment, which would mandate indifference (e.g., perhaps by flipping a coin in every Gun versus Box decision), and a principle of definite priority, which would require prioritizing the definite standpoint (i.e., always destroying the box).

From this perspective, if we imagine many worlds with Gun versus Box scenarios, a principle of equal treatment would result, *ex post*, in an identical number of bad outcomes as a principle of definite priority. If the number of deaths is the same in the long run, the objection goes, then neither rule is more reasonably rejectable than the other on these grounds. If this is correct, then

37 Further, under most views of reasons, nonexistent people cannot provide any reasons at all. See Martin, "Navigating Nonidentity," 97, especially n. 27.

38 As far as I am aware, Sheinman is the first to systematically put forward this distinction, arguing that principle contractualism is vulnerable to a charge of "principle worship" that is analogous to the "rule worship" objection in consequentialism ("Act and Principle Contractualism"). The possibility and nature of act contractualism has been the subject of recent discussion. For a defense of the possibility of act contractualism, for example, see Bourguignon, "On the Possibility of Act Contractualism." And for further discussion, see Salein, "Leaving Principle Contractualism Behind?" I return to Sheinman's discussion shortly.

39 Sheinman, "Act and Principle Contractualism"; and Bourguignon, "On the Possibility of Act Contractualism."

the equal treatment principle appears to be a natural consequence of rule-based contractualism in this context. As such, the earlier argument for rejecting the equal treatment principle is applicable only to act contractualism but not to rule contractualism.⁴⁰ This is a serious objection, as it threatens to undercut the applicability of this argument to the canonical and most widely discussed formulation of contractualism. However, I believe that the main argument put forward in section 2 works in both a rule-based and act-based understanding of contractualism.

To see why, consider the choice between the two competing principles, equal treatment and definite priority, across two levels of justification: the justifiability of the act relative to the principle (*justifiable*_{RULE}) and the justifiability of the act itself (*justifiable*_{ACT}). Even granting the premise of the objection—namely, that since both the equal treatment and definite priority principles produce the same number of deaths, they are both equally justifiable_{RULE}—there still is a significant difference in overall justifiability between equal treatment and definite priority. The difference is visible when we examine the justifiability of the acts licensed by each rule: the definite priority principle is unique because in Gun versus Box, it *always* licenses actions that are both justifiable_{RULE} and justifiable_{ACT}. Even in a tragic scenario where the box is full, the choice is still justifiable (both in act and rule terms) because it is a choice between two equally strong complaints. The principle of equal treatment, however, does not share this feature. Even granting that equal treatment may be justifiable_{RULE} in approximately one-sixth of cases, it licenses an action that is unjustifiable_{ACT}—namely, shooting the gun when the box is empty. I suggest that this asymmetry creates a clear reason to prefer the definite priority principle. A principle whose licensed actions are always justifiable at both the act level and the rule level is superior to a principle that, while licensing actions that are justifiable at the rule level, sometimes demands that agents perform actions that are unjustifiable at the act level. Put another way, to choose the equal treatment principle over the definite priority principle is to at least sometimes create a needless conflict between what the rule permits and what is justifiable to the actual person in front of us. As such, the definite priority principle dominates the equal treatment principle, given our uncertainty, as it avoids the risk of acting unjustifiably.⁴¹ In other words, the definite priority principle is sometimes better than

40 My thanks to a helpful anonymous reviewer for suggesting this objection and discussion.

41 Here, 'dominance' refers to a principle of choice where one option is considered superior if it is better in at least one possible state of affairs and not worse in any other. For a classic discussion, see Luce and Raiffa, *Games and Decisions*. For a contemporary accessible overview of the dominance principle, see Peterson, *An Introduction to Decision Theory*.

the equal treatment principle (when the box is empty) and never worse than the equal treatment principle (when the box is full).

That justifiable_{RULE} and justifiable_{ACT} can come apart and that sometimes a justifiable rule can license an unjustifiable action lead to the problem of “principle worship” for rule-based versions of contractualism.⁴² While Hanoch Sheinman develops this “principle worship” objection on the basis that principle contractualism sometimes implausibly “forego[es] unrejectability in the actual world for the sake of [principle] conformity,” this objection is also relevant to the relational aspect of contractualism that is central to my own argument.⁴³ The reason we care about finding unrejectable principles is because we value standing in a relationship of mutual recognition with other individuals. Given this, a contractualist rule or principle is only as good as the justifiability of the actions it prescribes. If a rule directs me to act in a way that is not justifiable to the person before me, the fact that the rule produces justifiable outcomes on average or in the abstract does not negate the justificatory failure with respect to that individual.

This points to a more fundamental feature of any plausible version of contractualism: the normative status of contractualist principles is contingent on the standpoints they affect actually being occupied. As such, the distinction between definite standpoints (which we know are occupied) and indefinite standpoints (which may be empty) remains relevant even when formulating and applying general principles, as in the standard version of the theory. As the earlier Gun versus Box choice demonstrates, the complaint from the definite standpoint in Gun has a different justificatory status than the complaint from the indefinite standpoint in Box. Any proposed rule (e.g., the equal treatment principle) must itself be justifiable to all. A rule that ignores the difference between definite and indefinite standpoints is reasonably rejectable by those in definite standpoints, as it requires them to bear risks for the sake of standpoints that may not be occupied. As such, any justifiable rule must take account of the distinction between definite and indefinite standpoints, making this distinction relevant even within the two-stage, rule-based version of the theory.

42 This is analogous to a well-known problem for rule utilitarianism. See Smart and Williams, *Utilitarianism*. For a clear recent discussion of this objection to rule-based ethical theories, see also Copp, “The Rule Worship and Idealization Objections Revisited and Resisted.” I leave further discussion of this aside for now, as I aim not to weigh in on the emerging debate of act versus rule versions of contractualism but only to suggest that we have reason to think that the problem raised here for the equal treatment principle also applies to rule-based versions of contractualism.

43 Sheinman, “Act and Principle Contractualism,” 302.

Finally, even if one were not convinced that this argument applies to the standard rule-based form of contractualism, this objection, at worst, limits the scope of my conclusion. Showing that *ex post* act contractualism cannot accept the equal treatment principle is itself still significant, even if it does not also hold for the standard rule-based version. It still clarifies the commitments of a major branch of contractualist thought and reveals a tension between the theory's relational foundations and an intuitively plausible principle of fairness. This contribution remains valuable even if it does not extend to the classic rule-based formulations of the theory and is relevant for comparing the overall plausibility of act versus rule formulations of the theory more generally.

4. WHY REJECTING EQUAL TREATMENT MATTERS: IDENTIFIED VICTIMS AND EX POST DISCOUNTING

I argue that contractualists should not endorse the equal treatment principle. Instead, in competitive cases with equal harms, contractualism must prioritize definite standpoints over indefinite standpoints. This has significant implications for other areas of contractualist theory, particularly as it relates to how contractualism addresses the identified victim bias. Further, it may provide a new and plausible basis for *ex post* forms of discounting.

As noted, this argument is also relevant to the debate over the normative status of the identified victim bias and questions over prioritizing identified versus statistical lives.⁴⁴ The debate over identified versus statistical lives centers on whether we should give priority to saving *identified* lives over *statistical* lives. To borrow a well-known example, consider whether we should rescue a single trapped miner or spend the same amount to implement safety measures that would save some number of unidentified future miners.⁴⁵ The first is an identified life, whereas the latter are statistical lives. People generally demonstrate a durable "bias" in favor of identified lives, and this bias sometimes appears in the context of ethics and public policy, e.g., in the form of "the rule of rescue."⁴⁶

44 For an excellent book covering various interdisciplinary aspects of the identified victim bias, see Cohen et al., *Identified Versus Statistical Lives*.

45 For discussion of the ethical aspects of this case, see Cohen et al., "Statistical Versus Identified Persons." For a recent grounding of the identified victim bias in the notion of decreasing marginal moral value of survival chances, see also Steffánsson, "Identified Person 'Bias' as Decreasing Marginal Value of Chances."

46 Here, I do not intend 'bias' to be used pejoratively. 'Identified victim effect' is sometimes used to avoid the negative connotations of the term 'bias', but I use 'bias' as it is still the most common way to refer to the phenomenon at hand. Regarding the identified victim bias, see Jenni and Loewenstein, "Explaining the Identifiable Victim Effect." For critical

While many argue that the identified victim bias is morally suspect, the contractualist rejection of the equal treatment principle provides a limited justification for prioritizing identified lives.⁴⁷ The distinction between identified and statistical lives often tracks a more fundamental distinction: the certainty we have about whether a given standpoint is or will be filled. When we face an identified victim, we know with certainty that there is an actual individual whose reasons we must consider. With statistical lives, however, we face uncertainty not just about outcomes but about the very existence of individuals who could provide personal reasons. This concern maps directly onto the distinction between Gun and Box. Just as the victim in Gun is a definite standpoint representing an actual individual, identified victims are represented by definite standpoints. Conversely, just as the person in the opaque box in Box may or may not exist to support reasons of the right type, statistical lives represent indefinite standpoints—we do not know if there actually are or will be individuals capable of grounding reasons.⁴⁸

This in turn provides a novel grounding for the identified victim bias. Rather than arguing that identifiability itself has moral significance, my argument here suggests that what matters is our certainty about the standpoints used to ground reasons for rejection. Identifiability serves as a proxy for this more fundamental moral consideration—namely, our certainty that there is an actual individual whose reasons we must consider.⁴⁹ When discussed this way, we can offer a similar argument to the previous section: prioritizing identified lives helps us to avoid the risk of unjustifiably sacrificing an actual person for an empty standpoint.

This reframing has important implications. First, it suggests the preference for identified lives is not necessarily irrational—at least, not for committed

discussion of the rule of rescue, see Brock and Wikler, “Ethical Challenges in Long-Term Funding for HIV/AIDS.”

47 E.g., Hope, “Rationing and Life-Saving Treatments.”

48 This aspect is often overlooked in discussion of the identified victim bias, as statistical outcomes are typically stipulated in advance. For extended criticism of this aspect of the literature and the way it may distort our ethical reasoning, see Fried, *Facing Up to Scarcity*.

49 In a previous paper, I attempt to give an *ex ante* contractualist argument for the identified victim bias (“An Uncertainty Argument for the Identified Victim Bias”). For criticism of my argument, see Gilbertson, “Indifference, Indeterminacy, and the Uncertainty Argument for Saving Identified Lives.” The justification presented in this article is not exactly comparable to the one I offered previously, as the current article focuses on an *ex post* version of contractualism. However, the justification for the bias is not vulnerable to the same objection that Gilbertson raises, as Gilbertson’s criticism focuses on the use of the *principle of indifference* and the challenges it introduces to *ex ante* contractualism. This current *ex post* justification does not turn on the principle of indifference in the same way.

contractualists—as it reflects a meaningful moral distinction about the certainty of standpoints (and thus, reasons for rejection). Second, it provides guidance about when this preference is justifiable: specifically, it is justifiable when identification tracks (the probability of) existence rather than mere salience. This aspect is sometimes obscured in discussions of the ethical aspects of the identified victim's bias, because the loss of statistical lives is usually stipulated to occur with certainty, thus guaranteeing that the standpoint is or will be filled. In other words, the framing of these cases often obscures the underlying features that make the divide between identified and statistical lives morally relevant in a contractualist framework.⁵⁰

This more nuanced understanding helps explain both the intuitive appeal (for those who have such an intuition) and the limits of the identified victim bias. It suggests we should in fact give special weight to identified lives but only insofar as identification helps us track the existence of actual individuals. Contractualism holds that what ultimately matters is actuality—the existence of real individuals whose complaints we must take account of—rather than identifiability *per se*. This provides a principled basis for when and why we should prioritize identified over statistical lives.

Beyond implications for the identified victim bias, and perhaps more significantly for contractualism, this argument may provide a basis for an account of *ex post* discounting.⁵¹ The argument here focuses on cases where expected losses are equal, as these are the types of cases that the equal treatment principle seeks to address. However, additional questions arise in cases where the burdens differ significantly between definite and indefinite standpoints (e.g., weighing a small burden for a definite standpoint versus a large burden for an indefinite one). Resolving these cases may require developing a form of *ex post* discounting based on standpoint uncertainty. Such discounting could operate on the likelihood of the standpoint being filled. For example, in Box, this means discounting the complaint of death by one-sixth, whereas in Gun, there would be no discounting, as we know that the potential victim of Gun exists, and we know with certainty that the victim-of-gun standpoint is occupied. So we would compare death for the victim in Gun against one-sixth chance of death for the person in the opaque box. This is not a matter of discounting the strength of the complaint itself as is the case in a traditional *ex ante* approach; it instead pertains to the likelihood of justification being possible. There are challenging questions about this approach, particularly regarding how theoretically

50 Again, for discussion regarding this stipulation of statistical outcomes, see Fried, *Facing Up to Scarcity*.

51 For discussion of an alternative approach to *ex post* discounting, see Steuwer, "Contractualism, Complaints, and Risk."

justified such a move may be. However, here I simply aim to highlight this possibility, as developing a full account of discounting based on standpoints rather than complaints requires a more extensive treatment than I am able to provide here.

This form of discounting could be a plausible way of developing *ex post* contractualism and, importantly, could also be employed with *ex ante* contractualism. While *ex post* discounting makes the strength of the complaint conditional on the likelihood of the relevant standpoint being filled, *ex ante* discounting would discount *both* the complaint and the standpoint. In other words, all *ex ante* probabilities would need to be conditional on the probability of the relevant standpoint being filled. However, this requires a fairly complicated redevelopment of contractualism, and meaningful discussion of this possibility is far beyond the scope of this article. In short, the distinction between definite and indefinite standpoints developed in this article may be able to support the development of a plausible form of discounting for *ex post* contractualism, based on how certain we are that a given standpoint is occupied.⁵² This is an important avenue for future research for the development of contractualist approaches to risk and uncertainty.

5. CONCLUSION

I have argued that contractualism cannot comply with the principle of equal treatment for equal statistical loss. The basic problem with the principle of equal treatment is that contractualism is a relational moral theory and, as such, is concerned primarily with *justifiability to actual individuals* rather than with outcomes themselves. In other words, what matters is not just expected statistical loss but whether the reasons for rejecting a given principle or action are grounded in actual moral relationships. This focus on justifiability to actual others requires contractualism to recognize a distinction between definite and indefinite standpoints. When we know a standpoint is occupied—when we are certain that at least one actual individual is affected—contractualism requires us to take that standpoint's reasons into account. I have called these standpoints definite standpoints. In contrast, indefinite standpoints involve uncertainty and represent instances when we are unsure if there is anyone to whom the standpoint corresponds (as in the person in the opaque box in the Box case). This matters because if no one actually occupies the standpoint, then there is no individual to whom justification is owed, and any reasons arising

52 However, it is unlikely to be applicable in many (perhaps even most) cases, as it is likely rather rare that we have explicit probabilities for standpoints being full, at least outside of philosophy thought experiments.

from such a standpoint cannot be considered in contractualist deliberation. If we were to act on the reasons of an empty standpoint to the detriment of an actual person, we would be acting in an unjustifiable manner. To hold otherwise would require that contractualism either recognizes reasons that have to do with something other than the moral relationship or holds that we have a moral relationship with standpoints themselves.⁵³

This distinction means that in some cases, we can have equal expected statistical losses that are not morally equal because the probability in one pertains to a certainly existing person facing a probabilistic harm, and the probability in the other pertains to the probability of whether there is a person who stands to be harmed at all. Although these two kinds of cases may be probabilistically equal and may entail equal expectations of statistical loss, they are not morally equal. This is because prioritizing the second risks acting unjustifiably in a way that prioritizing the first never does. If the indefinite standpoint turns out to be unoccupied, it cannot provide reasons of the kind demanded by contractualism. As such, acting on the basis of such reasons against the interests of an actual person is unjustifiable. Given that contractualism requires acting justifiably, it also requires choosing actions that are certainly justifiable over actions that are possibly or riskily justifiable. As a result, I have argued that we should prioritize definite over indefinite standpoints in competitive cases with equal expected burdens or benefits. Thus, to avoid acting unjustifiably, contractualist must at least sometimes prioritize definite over indefinite standpoints, which means that in at least some cases, contractualists must reject the equal treatment principle.

This argument has broader implications beyond the need to reject or revise the equal treatment principle. First, it provides a new, *ex post* contractualist justification for the identified victim bias. The preference for identified over statistical lives is frequently dismissed as irrational or otherwise without ethical justification, but the distinction between definite and indefinite standpoints and the contractualist requirement to treat them differently may provide a moral justification for the bias. It is not identifiability per se that matters but rather that identified lives involve reasons that we are certain we must consider. Second, this argument lays groundwork for a possible form of *ex post* discounting. If the normative status of a standpoint depends on whether it is occupied, then one way to approach certain cases of risk and uncertainty is to discount the reasons of indefinite standpoints relative to definite standpoints. This approach to discounting could apply to both *ex post* and *ex ante* versions of contractualism. Further work is needed to explore how exactly

53 Gibb, "Relational Contractualism and Future Persons."

such a discounting approach might be developed. Finally, while this article has focused exclusively on *ex post* contractualism, a full exploration of how the definite/indefinite standpoint distinction operates within an *ex ante* framework remains an important direction for future research.

Ultimately, what this argument shows is that equal expected statistical losses are not always morally equal. When we determine what is justifiable, we must look beyond expected statistical losses and consider the underlying moral relationships and how they influence the justifiability of inflicting such expected statistical losses. If contractualism aims to remain a theory that takes the moral relationship between individuals seriously, then in competitive cases, it must prioritize the interests of definite standpoints over indefinite standpoints, and thus, it must reject the equal treatment principle.⁵⁴

Jagiellonian University
jay@sdu.dk

REFERENCES

- Ashford, Elizabeth, and Tim Mulgan. "Contractualism." *Stanford Encyclopedia of Philosophy* (Summer 2018). <https://plato.stanford.edu/archives/sum2018/entries/contractualism/>.
- Bourguignon, Léa. "On the Possibility of Act Contractualism." *Australasian Journal of Philosophy* 102, no. 2 (2024): 1–19.
- Brock, Dan W., and Daniel Wikler. "Ethical Challenges in Long-Term Funding for HIV/AIDS." *Health Affairs* 28, no. 6 (2009): 1666–76.
- Cohen, I. Glenn, Norman Daniels, and Nir M. Eyal, eds. *Identified Versus Statistical Lives: An Interdisciplinary Perspective*. Population-Level Bioethics Series. Oxford University Press, 2015.
- . "Statistical Versus Identified Persons: An Introduction." In Cohen, Daniels, and Eyal, *Identified Versus Statistical Lives*.
- Copp, David. "The Rule Worship and Idealization Objections Revisited and Resisted." *Oxford Studies in Normative Ethics* 10 (2020): 131–55.

54 I would like to thank Callum MacRae and Christoph Merdes for valuable discussion of some of the ideas in this article, and Anastasiia Babash for helpful comments on an earlier draft of this article. This research is part of project no. 2022/47/P/HS1/02511, co-funded by the National Science Centre and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 945,339. For the purpose of open access, the author has applied a CC-BY public copyright license to any author accepted manuscript (AAM) version arising from this submission.

- Darwall, Stephen. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Harvard University Press, 2006.
- Eyal, Nir. "Concentrated Risk, the Coventry Blitz, Chamberlain's Cancer." In Cohen, Daniels, and Eyal, *Identified Versus Statistical Lives*.
- Frick, Johann. "Contractualism and Social Risk." *Philosophy and Public Affairs* 43, no. 3 (2015): 175–223.
- Fried, Barbara H. *Facing Up to Scarcity: The Logic and Limits of Nonconsequentialist Thought*. Oxford University Press, 2020.
- Gibb, Michael. "Relational Contractualism and Future Persons." *Journal of Moral Philosophy* 13, no. 2 (2016): 135–60.
- Gilbertson, Eric. "Indifference, Indeterminacy, and the Uncertainty Argument for Saving Identified Lives." *Journal of Applied Philosophy* 41, no. 3 (2024): 480–97.
- Gillies, Donald. *Philosophical Theories of Probability*. Routledge, 2006.
- Hayenhjelm, Madeleine, and Jonathan Wolff. "The Moral Problem of Risk Impositions: A Survey of the Literature." *European Journal of Philosophy* 20, no. 1 (2012): 26–51.
- Hope, Tony. "Rationing and Life-Saving Treatments: Should Identifiable Patients Have Higher Priority?" *Journal of Medical Ethics* 27, no. 3 (2001): 179–85.
- Jenni, Karen, and George Loewenstein. "Explaining the Identifiable Victim Effect." *Journal of Risk and Uncertainty* 14, no. 3 (1997): 235–57.
- John, Stephen David, and Emma J. Curran. "Costa, Cancer and Coronavirus: Contractualism as a Guide to the Ethics of Lockdown." *Journal of Medical Ethics* 48, no. 9 (2022): 643–50.
- Katz, Corey. "Contractualism, Person-Affecting Wrongness and the Non-Identity Problem." *Ethical Theory and Moral Practice* 21, no. 1 (2018): 103–19.
- Kim, Suzie. "On the Need for Real Dialogue: What's Wrong with Monological Contractualism?" *European Journal of Philosophy* 27, no. 4 (2019): 939–56.
- Kumar, Rahul. "Risking and Wronging." *Philosophy and Public Affairs* 43, no. 1 (2015): 27–51.
- Luce, R. Duncan, and Howard Raiffa. *Games and Decisions: Introductions and Critical Survey*. Wiley, 1957.
- Martin, Desa Valeska. "Navigating Nonidentity: Scanlonian Contractualism and Types of Persons." *Journal of Ethics and Social Philosophy* 29, no. 1 (2024): 86–106.
- Muñoz, Daniel. "Each Counts for One." *Philosophical Studies* 181, no. 10 (2024): 2737–54.
- Oberdiek, John. *Imposing Risk: A Normative Framework*. Oxford University Press, 2017.

- Otsuka, Michael. "Risking Life and Limb." In Cohen, Daniels, and Eyal, *Identified Versus Statistical Lives*.
- . "Scanlon and the Claims of the Many Versus the One." *Analysis* 60, no. 3 (2000): 288–93.
- Parfit, Derek. *On What Matters*, vol. 2. Oxford University Press, 2011.
- Peterson, Martin. *An Introduction to Decision Theory*. Cambridge University Press, 2017.
- Reibetanz, Sophia. "Contractualism and Aggregation." *Ethics* 108, no. 2 (1998): 296–311.
- Rüger, Korbinian. "On *Ex Ante* Contractualism." *Journal of Ethics and Social Philosophy* 13, no. 3 (2018): 240–59.
- Salain, Valentin. "Leaving Principle Contractualism Behind? A Response to Salomon." *Journal of Ethics and Social Philosophy* 30, no. 1 (2025): 146–54.
- Saunders, Ben. "A Defence of Weighted Lotteries in Life Saving Cases." *Ethical Theory and Moral Practice* 12, no. 3 (2009): 279–90.
- Scanlon, T. M. *What We Owe to Each Other*. Belknap Press, 1998.
- Sheinman, Hanoch. "Act and Principle Contractualism." *Utilitas* 23, no. 3 (2011): 288–315.
- Smart, John Jamieson Carswell, and Bernard Williams. *Utilitarianism: For and Against*. Cambridge University Press, 1973.
- Southwood, Nicholas. *Contractualism and the Foundations of Morality*. Oxford University Press, 2013.
- . "Moral Contractualism." *Philosophy Compass* 4, no. 6 (2009): 926–37.
- Spiekermann, Kai. "Good Reasons for Losers: Lottery Justification and Social Risk." *Economics and Philosophy* 38, no. 1 (2022): 108–31.
- Stefánsson, H. Orri. "Identified Person 'Bias' as Decreasing Marginal Value of Chances." *Noûs* 58, no. 2 (2024): 536–61.
- Steuwer, Bastian. "Contractualism, Complaints, and Risk." *Journal of Ethics and Social Philosophy* 19, no. 2 (2021): 111–39.
- Suikkanen, Jussi. "*Ex Ante* and *Ex Post* Contractualism: A Synthesis." *Journal of Ethics*, 23, nos. 3–4 (2019): 391–410.
- Tadros, Victor. "Controlling Risk." In *Prevention and the Limits of the Criminal Law*, edited by Andrew Ashworth, Lucia Zedner, and Patrick Tomlin. Oxford University Press, 2013.
- Taurek, John M. "Should the Numbers Count?" *Philosophy and Public Affairs* 6, no. 4 (1977): 293–316.
- Wasserman, David. "Let Them Eat Chances: Probability and Distributive Justice." *Economics and Philosophy* 12, no. 1 (1996): 29–49.
- Zameska, Jay. "An Uncertainty Argument for the Identified Victim Bias." *Journal of Applied Philosophy* 39, no. 3 (2022): 504–18.

JOURNAL of ETHICS & SOCIAL PHILOSOPHY
<http://www.jesp.org>
ISSN 1559-3061

The *Journal of Ethics and Social Philosophy* (JESP) is a peer-reviewed online journal in moral, social, political, and legal philosophy. The journal is founded on the principle of publisher-funded open access. There are no publication fees for authors, and public access to articles is free of charge. Articles are typically published under the CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL-NODERIVATIVES 4.0 license, though authors can request a different Creative Commons license if one is required for funding purposes.



Funding for the journal has been made possible through the generous commitment of the Division of Arts and Humanities at New York University Abu Dhabi.

جامعة نيويورك أبوظبي



NYU ABU DHABI