JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY

VOLUME XXVI · NUMBER 3
February 2024

ARTICLES

How to Read a Riot
Ricky Mouser

Dismissing Blame
Justin Snedegar

The Problem of Basic Equality:
A Constructive Critique
Nikolas Kirby

What Time Travel Teaches Us about Moral Responsibility
Taylor W. Cyr and Neal A. Tognazzini

Paternalism and Exclusion
Kyle van Oosterum

Maxim and Principle Contractualism
Aaron Salomon

Nonnaturalism, the Supervenience Challenge, Higher-Order Properties, and Trope Theory
Jussi Suikkanen

DISCUSSIONS

Privileged Citizens and the Right to Riot:
A Reply to Pasternak
Thomas Carnes

Gaslighting and Peer Disagreement
Scott Hill

OPEN ACCESS FREE FOR AUTHORS AND READERS ISSN 1559-3061

JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY

VOLUME XXVI · NUMBER 3
February 2024

ARTICLES

445 How to Read a Riot
Ricky Mouser

469 Dismissing Blame
Justin Snedegar

495 The Problem of Basic Equality:
A Constructive Critique
Nikolas Kirby

523 What Time Travel Teaches Us about Moral Responsibility
Taylor W. Cyr and Neal A. Tognazzini

547 Paternalism and Exclusion
Kyle van Oosterum

571 Maxim and Principle Contractualism
Aaron Salomon

601 Nonnaturalism, the Supervenience Challenge, Higher-Order Properties, and Trope Theory
Jussi Suikkanen

DISCUSSIONS

633 Privileged Citizens and the Right to Riot:
A Reply to Pasternak
Thomas Carnes

641 Gaslighting and Peer Disagreement
Scott Hill

JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY
http://www.jesp.org

The *Journal of Ethics and Social Philosophy* (ISSN 1559-3061) 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 and is available to all readers under the CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL-NODERIVATIVES 4.0 license. 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.

The *Journal of Ethics and Social Philosophy* 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 associate 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.

Journal of Ethics and Social Philosophy
VOL. 26, NO. 3 · FEBRUARY 2024
https://doi.org/10.26556/jesp.v26i3.2316

HOW TO READ A RIOT

Ricky Mouser

GEORGE FLOYD, a 46-year-old Black man, was murdered by Derek Chauvin, a White police officer, in Minneapolis, Minnesota, on May 25, 2020.[1] Although Chauvin was captured on video kneeling on Floyd's neck for around eight minutes, his official police report grossly misrepresented the nature of their encounter. In response, thousands of peaceful protestors gathered in the streets and marched on the Third Precinct police headquarters.[2]

Later that night, after the larger crowd had disbanded, a few hundred protestors threw rocks and water bottles at the building and began smashing police car windows in the parking lot. Police overreacted by firing tear gas and rubber bullets into the crowd, escalating and inciting further violence.[3] And in the days that followed, militarized police continued to meet protests about their brutality with overwhelming force. While many protestors remained unambiguously peaceful, others blocked highways, set Chauvin's police station ablaze, and (infamously) looted a Target. As protests spread from Minneapolis across the nation, after-action evaluations in city after city consistently confirmed that police deployed and provoked violence irresponsibly.[4]

Yet the immediate response from elected officials at every level was remarkably uniform. Minneapolis Mayor Jacob Frey lamented that "what started as largely peaceful protests for George Floyd have turned to outright looting and domestic terrorism in our region." Minnesota Governor Tim Walz activated the National Guard, proclaiming that "the situation in Minneapolis is no longer, in any way, about the murder of George Floyd. It is about attacking civil society, instilling fear, and disrupting our great cities." And on Twitter, President Trump warned: "When the looting starts, the shooting starts."[5]

1 Taylor, "George Floyd Protests."

2 Kaul, "Seven Days in Minneapolis."

3 Caputo, Craft, and Gilbert, "'The Precinct Is on Fire.'"

4 Barker, Baker, and Watkins, "In City after City, Police Mishandled Black Lives Matter Protests."

5 Taylor, "George Floyd Protests."

Comments such as these make plain a variety of presuppositions: that political protests are legitimate only insofar as they are peaceful; that when they turn violent, they become a menace to society itself; and that in response, overwhelming state violence is justified. In these discussions, the word "riot" itself often comes to be used as an epithet, and the very idea that political violence might count as a legitimate form of *protest* is gravely contested. In response, I offer a radical reassessment of political rioting as a deeply expressive rejection of the political status quo.[6] I argue that political rioting, as "the language of the unheard," can be a proportionate, minimally harmful means of directing the attention of the state and the broader public toward urgent structural injustices.[7] Along the way, I situate political rioting between civil disobedience and political revolution to highlight its unique expressive force.

In section 1, I note that political rioting goes beyond civil disobedience by openly contesting the value or applicability of civility under the political status quo. In section 2, I argue for reading the immediate aims of political rioting as fundamentally opposed to those of political revolution. Where political revolutionaries aim at separation from the state, political rioters paradigmatically desire more full inclusion within it. In section 3, I build on Aria Pasternak's innovative interpretation of political rioting as a defensive harm while highlighting its function as a publicly *expressive* form of protest. I leverage this insight in section 4 to expand the controversial "success constraint" on defensive harm to include not only material but also *existential*, fundamentally respect-based harms.

1. CIVIL DISOBEDIENCE AND POLITICAL RIOTING

What counts as specifically political rioting, as opposed to the sort of rioting that sometimes accompanies sporting events? I argue that political rioters do not break the law "mindlessly" or for merely selfish gain but to (as I will call it) *ocularize*, or render spectacularly visible, some purported injustice. In other words, we should read political rioting as a form of principled lawbreaking.[8]

To render this plausible, consider political rioting in relation to civil disobedience, a form of principled lawbreaking usually considered more acceptable on its face. Paradigmatic images of each are vivid enough: civilly disobedient protestors stage lunch-counter sit-ins, spoil draft cards with blood, and (more

6 I say "expressive," not "communicative," as successful expression does not require uptake.

7 Martin Luther King Jr., as quoted in Rothman, "What Martin Luther King Jr. Really Thought about Riots."

8 I allow for the conceptual possibility of principled lawbreaking in support of morally wrong causes, such as apartheid. The principles at work in such a case are simply heinous, making such rioting *worse* than so-called random violence.

or less) cooperate when officers come to arrest them. On the other hand, political rioters smash windows, torch cars, and physically resist police intervention. How can we bring these clusters of images together to form the basis of a useful conceptual distinction?

Candice Delmas analyzes civil disobedience as

> a principled and deliberate breach of law intended to protest unjust laws, policies, institutions, or practices, and undertaken by agents broadly committed to basic norms of civility. This means the action is public, non-evasive, nonviolent, and broadly respectful or civil (in accordance with decorum).[9]

Delmas conceives of civility as "decorum" concerning "the ways citizens ought to interact with each other in the public sphere, when debating political questions."[10] Building on this, civility involves ocularizing one's own deference to discourse-relevant norms, usually long before implied threats of enforcement need to become actualized or even too openly visible.[11] This requires visible cooperation with and anticipation of the public's and the state's responses to one's actions.[12]

Of course, this is what makes civil disobedience so striking—even while breaking laws they deem unjust, civil disobedients *otherwise* ocularize their civility, for moral or other reasons.[13] Thus, I distinguish between civilly disobedient protestors and political rioters at the level of tactics, in terms of how they ocularize the purported injustice they are protesting.[14] After breaking the

9 Delmas, *A Duty to Resist*, 17.

10 Delmas, *A Duty to Resist*, 43.

11 Compare William E. Scheuerman's notion of civility "as shared commitment to a common political project" ("Why Not Uncivil Disobedience?," 11).

12 For simplicity, I will speak of principled lawbreaking in relation to states, but there is no reason in principle barring corporations or other bodies from being resisted by these or similar means.

13 Note that this does not mean that under civil disobedience civility must be *total*. But civil disobedience aims to be read as cooperative (to make its motivating concerns more legible) in a way that political rioting spurns.

14 Candice Delmas argues convincingly that American civil rights groups in the 1960s "adopted their particular style of civil disobedience for context-dependent, tactical purposes. Yet theorists and pundits turned these tactics into deep moral commitments on the part of agents supposedly eager to demonstrate their endorsement of the state's legitimacy, and placed these subjective requirements at the core of their defense of real-world civil disobedience" (*A Duty to Resist*, 27–28). While I do not deny that many protestors prefer civil disobedience for moral reasons, my analysis emphasizes the ultimately tactical and noncategorical basis of the distinction between civil disobedience and other forms of protest. See also Cobb, *This Nonviolent Stuff'll Get You Killed*, 8.

law, principled lawbreakers can react more or less cooperatively to the state's response. And the distinction falls out of this: that is, civil disobedience involves breaking the law and then otherwise ocularizing one's own *relative* civility in interacting with the state's intervention of "law and order," thereby expressing confidence that justice can be achieved by means of procedural cooperation with state institutions, either directly (by these institutions' just operation) or indirectly (by leveraging outrage at their unjust operation). On the other hand, political rioting involves breaking the law and then ocularizing one's own violent rejection of civility in interacting with these same state response mechanisms, expressively contesting the appropriateness of even the appearance of procedural cooperation with the state given circumstances on the ground.[15]

In making a show of cooperating with the state's responses to them, civil disobedients are *tactically civil*, accepting (at least outwardly) the legitimacy of civil procedure under the political status quo and upholding civility as a genuine civic virtue. In this way, they leverage the symbolic imagery of the state arresting and punishing dissenters who are otherwise visibly cooperative with the state, daring the state to perform mass arrests, overcrowd prisons, defend challenged laws in court, and so on.[16] This desire to ocularize their civility often (though not always) leads civil disobedients to eschew physical violence altogether.

Even so, note that on my analysis of civil disobedience, the lawbreaking act itself may be covert, evasive, offensive, or even violent, so long as its subsequent ocularization is not. Imagine a case where, to protest laws criminalizing rioting, a group of protestors with civilly disobedient principles *riots* in the streets (to pointedly break the law) and then, when the police show up, dutifully turns itself in to the authorities for arrest and processing. It would be true that this

15 Compare Thomas E. Hill Jr.'s discussion of *disassociation* from evil ("Symbolic Protest and Calculated Silence," 90–95).

16 John Rawls argued that willingness to *accept* legal punishment was necessary for civil disobedience to function as an effective mode of address to the majority holding political power while still showing fidelity to law (*A Theory of Justice*, 366). But many contemporary theorists find the requirement of submission to legal punishment unduly restrictive. Piero Moraro argues that because there are other ways that a civil disobedient can be answerable to their fellow citizens for their legal wrongdoing besides accepting punishment, we should recognize a justificatory gap between "breaching the law" and "being liable to punishment" ("On (Not) Accepting the Punishment for Civil Disobedience," 509). The underlying problem, Erin Pineda suggests, is that these "liberal and deliberative accounts" of civil disobedience misinterpret *all* civil disobedience "as either oriented toward moral suasion or as modestly reformist," overlooking the possibilities of more radical social disruption and solidarity building ("Civil Disobedience and Punishment," 20–21). These concerns extend to political rioters, who may also be accountable to their fellow citizens in other ways besides willingly accepting (often disproportionate, example-making) punishments.

group performed an act of civil disobedience by acting violently, at least initially, only ocularizing their civility afterward. Similarly, I think civil disobedients might count among their numbers those who strip publicly for political purposes and even some whistleblowers. Civil disobedience need not be as staid as we might commonly imagine, so long as it ultimately ocularizes civility.[17]

On the other hand, political rioters make a show of publicly breaking a whole slew of laws to create a zone of pointed lawlessness and then refusing to cooperate with the state's responses to them. They are *tactically uncivil*, at pains to ocularize their rejection of civil procedure and its ease and comfort with the political status quo. In this way, they leverage the symbolic imagery of pointedly existing "outside" the purview of the state's authority, if only for a moment, and then daring to resist the state's predictable physical reassertion of control over them. They cast the state as a militarized invader of its own public spaces, whose citizens resist the arrival of its mechanisms of "law and order" as unwelcome via what (expressively) approaches a miniaturized domestic war. This desire to ocularize their uncooperativeness is why political rioters are usually eager to employ particularly visible measures of physical violence against property, such as torching buildings and overturning cars.[18]

17 Of course, cooperation may be not just imprudent but morally corrosive. Edward Snowden would render himself complicit in grave political injustices were he to engage with the secretive kangaroo courts that await him in the United States. Scheuerman notes that "accepting penalties only makes sense if disobedients can count on legal proceedings embodying basic legal virtues" (*Civil Disobedience*, 132–33).

18 *Need* political rioters be violent? Stephen D'Arcy emphasizes civil defiance instead, defining a *riot* as "an outbreak of civil defiance, in which a crowd openly, directly, and persistently rejects the authority of the established legal order and its enforcers in the military or the police" (*Languages of the Unheard*, 145). Says D'Arcy, "I join the historians in treating violence, or harm to persons or property, as a nonessential feature of rioting. It is quite possible to join in a riot and participate fully in it, without acting to harm any person or damage any property" (*Languages of the Unheard*, 146). Other theorists disagree. Jonathan Havercroft understands the riot as a self-organizing crowd that disrupts the state's monopoly on violence and breaks laws concerning public assembly to express grievances outside of normal political processes ("Why Is There No Just Riot Theory?," 911). And Avia Pasternak insists that "political rioters resort to spontaneous, disorganized, public collective violence in order to protest against and to defy their political order" ("Political Rioting," 385).

Surely riots must use or at least threaten violence; after all, Mahatma Gandhi's huge crowds were civilly defiant but certainly not riots. But D'Arcy's emphasis on defiance foregrounds the expressive nature of riots, which *explains* the political riot's violence. Havercroft argues that "the British Crown … invented the concept of rioting as a crime in order to set limits on protest and dissent" ("Why Is There No Just Riot Theory?," 918). Instead of centering political rioting's violence, as law enforcement does to justify brutal crackdowns, we should read political riots as "mass rejections of constituted legal authority" that *use* (or threaten) violence to protest beyond the boundaries of civility (D'Arcy, *Languages of the Unheard*, 146).

Note that both civil disobedients and political rioters take on a great deal of personal risk, although some political rioters may take on even more risk by challenging law enforcement's capacity for physical violence head-on, potentially meeting force with force. But this additional risk brings certain advantages. It can render the standing threat of state violence more plainly visible, in the form of open conflict in the streets between militarized police and ordinary citizens.[19] Additionally, it expresses a daring willingness on the part of political rioters to subject themselves to these additional risks, demonstrating just how illegitimate they take the state mechanisms of "law and order" to be. By leveraging physical violence, political rioters openly announce that the norms of civility do not or should not apply given their *current* relationship with the state.

We already have quite a bit on the table, so allow me to summarize. I understand civil disobedience and political rioting as two forms of principled lawbreaking. They are distinguished by whether, apart from the lawbreaking act itself, participants ocularize their civility (roughly, compliance with discourse-relevant norms) or violently ocularize their incivility (noncompliance). This is ultimately a difference in *tactics* with significant expressive upshot.

2. POLITICAL RIOTING AND POLITICAL REVOLUTION

Define *political revolution* as group political action with the end of separation from the state. (I have in mind both violent revolutions, such as the American Revolution, and nonviolent revolutions, such as the Indian independence movement under Mahatma Gandhi.) In this section, I argue that unlike political revolutionaries, who aim at separation from the state, political rioters paradigmatically desire more full inclusion within it.[20]

Avia Pasternak notes that while defensive wars are fought against foreign aggressors, political rioters clash with their own state.[21] But at least in the case of many citizens, this difference is largely nominal. Even modern liberal democracies regularly fail to ensure that all citizens' basic liberties are institutionally

19 Admittedly, this benefit comes with corresponding costs in terms of *expressive clarity*. For some onlookers, the injustice being protested may be obscured by the public violence being used to ocularize it. Many seem to view any form of public violence as *ipso facto* unjustified and harbor double standards regarding the use of state violence. I do not think these reflect a common *considered* view—few condemn all public violence in principle—but it is a distinct barrier that political riots face when their grievances are not yet widely understood to be deep enough to merit violent response.

20 Even so, to the extent that political rioting harnesses the material threat of political revolution, it operates with the *seed* of revolutionary violence already in hand.

21 Pasternak, "Political Rioting," 386.

guaranteed in practice, either by inaction or active violation on the state's part. Call the resulting marginalized segments of the citizenry *deeply aggrieved*.[22] In many cases, cries of injustice and efforts toward institutional reform by deeply aggrieved citizens have borne little fruit for decades or generations on end, leaving these citizens on the outside of the state looking in.

I submit that the political rioters who came out to protest the murder of George Floyd either were themselves deeply aggrieved citizens or were protesting on behalf of such citizens. Police in America, acting as the state-sponsored arm of the law, have continued to brutalize Black citizens for generations. Katie Nodjimbadem notes just how *little* has changed since Martin Luther King Jr.'s insistence in his famous 1963 "I Have a Dream" speech that "we can never be satisfied as long as the Negro is the victim of the unspeakable horrors of police brutality."[23] Life under such a state is so precarious for Black citizens, Stephan Schwartz argues, that it is often the same as life under a foreign occupier in all but name:

> The truth that almost none of us who are White get is that 57 years after Martin Luther King's I Have a Dream speech, 56 years after the Civil Rights act of 1964, and 55 years after the Voting Rights Act of 1965, if you are Black or Brown, and particularly if you are a young Black man, for you America is like living in an occupied country where any interaction with the police is to be avoided.[24]

I take this charge of foreign occupation very seriously. For the sake of argument, let us grant that citizens have *pro tanto* duties to observe the sovereignty of their state and abide by its laws.[25] But if civil disobedience can ever be justified, then the duty to abide by the state's laws must be defeasible when these laws are seriously unjust. Similarly, if political revolution can ever be justified, the duty to observe the sovereignty of one's state must be defeasible if the state does not

22 Of course, the same individual citizen may be deeply aggrieved along multiple intersectional axes.

23 Nodjimbadem, "The Long, Painful History of Police Brutality in the U.S."

24 Schwartz, "Police Brutality and Racism in America," 282.

25 Even this much is seriously contentious; various theorists have argued that *there is no such duty*, at least under circumstances of severe injustice. Delmas argues that "the very grounds supporting a duty to obey also impose duties to disobey under conditions of injustice" (*A Duty to Resist*, 8–9). Ten-Herng Lai thinks that "disobeying the law may be the best way of realizing the substantive or procedural values that underpin the duty to obey the law" ("Justifying Uncivil Disobedience," 90). David Lyons even argues that "the assumption of political obligation is morally untenable" in general ("Moral Judgment, Historical Reality, and Civil Disobedience," 31). These theorists may be right, but my aim is to grant as much as possible to my interlocutor without minimizing the gravity of actual systemic injustices.

systematically guarantee the basic liberties of its citizens.[26] Thus, a state that does not safeguard any of its citizens' basic liberties is no legitimate state at all; a state that does little to uphold many of its citizens' basic liberties is, at least for those citizens, hardly *their* state.

On this account, the respect that a state commands is ultimately derivative of and conditional on it showing respect to its citizens. Thus, for as long as the state causes or remains deaf to pervasive, systematic violations of basic liberties, or even while it drags its feet in addressing them, it not only renders itself alien to its deeply aggrieved citizens; it also expresses that these citizens are alien *to the state itself*, that they are not full citizens.[27] By failing to acknowledge the full rights of all of its citizens, the state thereby releases those same citizens from any legitimate duties to recognize the full sovereignty of the state.[28] And then how can the state complain when its citizens are uncivil when they refuse to ocularize compliance with the state's mechanisms of justice? This is not fully *their* state, and these are not truly *their* mechanisms of justice, by the state's own implicit admission.

To lay out some of the political responses available to deeply aggrieved citizens, I categorize various forms of principled lawbreaking below, based on their political means and ends.

Table 1. Principled Lawbreaking: Selected Means and Ends

	Tactical civility *(procedurally cooperative means)*	Extreme tactical incivility *(violently uncooperative means)*
Political protest *(end of reform)*	Protestive civil disobedience (Martin Luther King Jr.)	Protestive physical violence (political rioters)
Political revolution *(end of separation)*	Revolutionary civil disobedience (Mahatma Gandhi)	Revolutionary physical violence (George Washington)

How does political rioting, as the paradigmatic form of protestive physical violence, fit into this picture? Like protestive civil disobedience, political rioting involves protest of the state with the (often implicit) end of reform—a political

26 This notion of reciprocally supporting rights has roots in Rodin's "Justifying Harm" (77).

27 Consider Mike Pence's revealing proclamation to the 2020 Republican National Conference that "the American people know we do not have to choose between supporting law enforcement and standing with our African-American neighbors to improve the quality of their lives, education, jobs and safety" (Epstein, "Full Transcript").

28 Defensive harms may be committed on another's behalf, so I am comfortable with the claim that the state's maintaining some of its citizens in a state of deep aggrievement is at least potentially sufficient for *all* of its citizens to find the state's claims of sovereignty over them damaged. In this way, "injustice anywhere is a threat to justice everywhere" (King, "Letter from a Birmingham Jail").

riot shouts "Things must change!" And like revolutionary physical violence, political rioting pursues its aims by physically violent means. But political rioters are not revolutionary soldiers; they aim not to separate themselves permanently from the state by force but to make themselves heard, if not by the state, then at least by the public and, perhaps most centrally, by themselves.[29] Reading political rioting this way, as an expressive protest demanding more full inclusion within the state, lets us understand political rioting as an ultimately conciliatory effort to redemocratize protestors' relationship with "their" state.[30]

Of course, few groups act as uniformly as this idealization suggests. As a general point, large-scale group actions are rarely univocal, and they need not be to be justified in whole or in part. Differences in individual intentions affect both the ends and means that political actors adopt. (For example, a single group event could involve both civil disobedience and political rioting if one part of the group ocularizes civility while another ocularizes incivility.) This is why we should prefer a generalized approach for justifying principled lawbreaking that can abstract away from at least some of these differences. In the next section, I attempt to provide just that.

3. Principled Lawbreaking as a Defensive Harm

Many might accept my line of argument thus far but hold that the political rioter's means are too dangerous, or their ends too ill-defined, to be justified. My aim in this section is to undermine these intuitions by drawing out the deep continuities between the means of political rioting and political revolution, as well as the ends of political rioting and protestive civil disobedience. I do this by building upon Avia Pasternak's work to argue that political rioting can (sometimes) be justified as a form of *defensive harm*—a term of art lifted from the philosophical literatures on self-defense and contemporary just war theory.

Per David Rodin, a defensive harm is a harm inflicted in order to avert or ameliorate harm to oneself or others.[31] We most often think of defensive harms in terms of physical violence: I punch a mugger to protect myself from harm, or a sovereign nation fights a defensive war to repel invaders. But nonphysically violent defensive harms are also possible, as when I copy human resources on

29 Compare the emphasis on *user* interpretation of value-based protest slogans in Myisha Cherry's "Value-Based Protest Slogans."

30 For more on the idea that principled lawbreaking might play an essential democratizing role, counteracting the ossification of the state, see Celikates, "Democratizing Civil Disobedience."

31 Rodin, "Justifying Harm," 74.

an aggressive email from a coworker or my union strikes in response to worsening labor conditions.

To be morally justified, defensive harms are subject to various constraints. Following Pasternak, I will focus on the *necessity constraint*, the *success constraint*, and the *proportionality constraint* in turn.[32]

First, the necessity constraint: roughly, a defensive harm is only permissible if it is the least harmful option that would still be efficacious.[33] To illustrate, if reasonable negotiation could prevent a neighboring country from invading, there would be no need to rush off and start a war instead. That level of force would be unjustified because open war would be completely unnecessary to avert the harms of a coercive military invasion, even if it would be successful.

Second, the success constraint: roughly, a defensive harm must have a reasonable chance of averting (or at least ameliorating) harm.[34] Where the necessity constraint calls on us to find the least harmful option that would still be efficacious, the success constraint insists that a permissible justified harm must still have a reasonable chance of attaining efficacy in the first place. For example, if fighting back against an irresistibly superior invading military force has no reasonable prospect of success, the thought is that it seems unwise to compound the misery of an inevitable invasion with the pointless slaughter of one's own forces. (I will return to this sort of example with a more critical eye in section 4.)

Finally, the proportionality constraint: roughly, a defensive harm must be proportionate to the harm it aims to avert.[35] It may be questioned how much independent work is left for the proportionality constraint to do, given that, between the necessity and success constraints, we already require the minimum amount of defensive harm that would still have a reasonable chance of success. But there could be cases where this amount of harm would be wildly disproportionate. Consider facing another nation's irresistible invading army. If the minimum efficacious defensive harm would be a massive nuclear

32 Pasternak, "Political Rioting," 386.

33 There are immediate complications relating to the full distribution of possible outcomes, our epistemic handle on these facts, and so on. There are also objective and subjective interpretations of this constraint: we might say that the necessity constraint requires that actors *actually* choose the least harmful option that is still efficacious (the objective interpretation), or we might require that actors *reasonably believe* that they do so (the subjective interpretation). See Lazar, "Necessity in Self-Defense and War." For simplicity, I will analyze all three constraints in objective terms.

34 Difficulties attend to formulating this constraint as well. See Statman, "On the Success Condition for Legitimate Self-Defense."

35 "Proportionate" need not mean "equal"; it is generally accepted that the defensive harm may *reasonably* outstrip the magnitude of the harm incurred. See Hurka, "Proportionality in the Morality of War."

strike obliterating the entire invading nation, this level of force would clearly be impermissibly disproportionate. Crucially, much of this force would be directed against civilians who would not be directly responsible for the invasion, or responsible with various partially excusing conditions, and so on.

The proportionality constraint, then, largely functions to govern the allocation of defensive harm to the responsible parties. Consider Jeff McMahan's distinction between narrow and wide proportionality: where judgments of *narrow* proportionality involve those who are liable for the harm to be avoided, judgments of *wide* proportionality involve, effectively, bystanders who are not liable in this way.[36] This framing clarifies our judgments about the nuclear strike, where the obliteration of innocent civilians is incredibly disproportionate in this wider sense, even though the narrower obliteration of the conniving enemy general would not be.

Crucially, for the defensive harm analysis to succeed, any political action must be justified not only in light of how it is executed but also why.[37] The necessity constraint covers the "why." For a political action to be justified as a *defensive* harm, it must be the least harmful option that would still have a reasonable chance of averting or ameliorating some harm. *A fortiori*, there must be some genuine harm targeted by the intended defensive harm for the constraint to be met.[38] This is why even a peaceful sit-in protesting the 2020 presidential election results could not be justified as a defensive harm.[39] Morally speaking, your political ends matter regardless of how polite you are.

36 McMahan, *Killing in War*, 20–21. Rodin builds on this to argue that narrow proportionality corresponds to liability justifications for defensive harm, whereas wide proportionality corresponds to lesser-evil justifications for defensive harm ("Justifying Harm," 78). Within riots that are justified overall, some innocent bystanders might still be harmed unavoidably. We may need to turn to lesser-evil justifications in these cases. Lesser-evil justifications assume that some evil is unavoidable and we (tragically) must choose the lesser evil. But crucially, lesser-evil justifications *do not* obviate the need to make amends. Thus, rioters may still be liable for harming innocent bystanders. Compare David K. Chan, *Beyond Just War*, 67, with Rodin, "Justifying Harm," 86.

37 Here I follow Delmas's suggestion that "basic human interests constrain both the legitimate goals and the appropriate means of resistance" (*A Duty to Resist*, 49).

38 Again, I am speaking in objective terms for simplicity. But even if we say that the action itself is objectively unjustified, questions of whether and to what extent individual protestors are blameworthy are separate. I sideline these difficult issues here.

39 The fact that the January 6, 2021, march on the Capitol amounted to an incipient *un*justified revolutionary effort only further worsens its moral standing. See Tavernise and Rosenberg, "These Are the Rioters Who Stormed the Nation's Capitol." Here my account is somewhat revisionary, though I think calling these political actors "rioters" (rather than, say, "insurrectionists") dramatically underrepresents their true aims simply because they were unsuccessful. I would find it rather bizarre to call the storming of the Bastille a mere *riot*.

On the other hand, to be justified as a defensive *harm*, we want to know the tactical considerations on the ground. Together, all three constraints require us to find means that will be suitably proportionate (and minimally harmful) while maintaining reasonable hope for success. Our political means are also morally relevant.

So, when might civil disobedience, political revolution, or political rioting be morally justified? In the end, the necessity constraint does most of the work charting the course of justified political action.[40] It is not difficult for things to be bad enough to warrant civilly disobedient protest; it is quite difficult for things to be so bad as to warrant political revolution. Within the chasm of space between these extremes, political rioting is at least sometimes the least harmful option that still has reasonable, proportionate prospects of success.

Unsurprisingly, civilly disobedient protest is the easiest of these three actions to justify. Under the necessity constraint, it is very often the case that lesser forms of law-abiding action or protest might not generate enough expressive force to achieve their intended political aims. But we know from experience that civilly disobedient tactics at least sometimes lead to real change. Under the success constraint, civilly disobedient protestors are often treated with a presumption of civic nobility or uprightness, which increases their odds of success by lending credibility to their political goals. Earned or not, this moralized reputation is tactically useful. And under the proportionality constraint, civil disobedience involves breaking the laws of the state, directly targeting the state under narrow proportionality. Of course, the effects of this principled lawbreaking may harm others under wide proportionality, but these are effects to be weighed. (On these sorts of grounds, ambulance drivers might be slower to strike illegally than construction workers.)

On the other hand, political revolution is much more difficult to justify. Under the necessity constraint, it is only occasionally the case that the minimal efficacious harm will involve trying to separate from the state rather than reform it. Under the success constraint, political revolutions are quite unlikely to succeed, particularly in highly militarized modern liberal democracies. And under the proportionality constraint, it is very hard to avoid the fallout of a political revolution causing indiscriminately wide harms, including against revolutionaries themselves. Political revolutions are dangerous, difficult undertakings whose effects are hard to foresee; even so, we still think they can be justified under extreme enough conditions. Indeed, we share a deep conviction that

40 Even so, all three constraints are required for political rioting to be justified. Further, if some constraints are unsatisfied, there is still the possibility that a riot is unjustified but excusable if rioters' actions remain reasonable given their circumstances.

political revolutions of centuries past secured the universal basic liberties that "we" now (at least nominally) enjoy within our modern liberal democracies.

So when might political rioting be justified? Pasternak argues that "even in democratic societies spontaneous violent protest can become the *only means* available for oppressed citizens to secure a range of valuable political goals," including changes in public policy, resistance to marginalization, and expression of angry defiance toward political authorities.[41] This is an appeal to the necessity constraint. The claim is: sometimes, a bit of uncivil resistance in the form of political rioting is needed to grab the attention of the state, the public, and even deeply aggrieved citizens themselves and potentially spark changes in how their ongoing oppression is understood and addressed. Cameras show up for fires and broken windows.[42] Note that in the United States, decades of legal action and lesser forms of protest, punctuated by very occasional riots, have not put an end to police brutality yet. Doing what we have been doing has done little to curb police brutality so far.[43]

Per the success constraint, there is some chance that political rioting will achieve concrete policy changes, but more importantly, it may change the tenor and focus of the interlocking conversations involving deeply aggrieved citizens, the broader public, and the state. In the aftermath of the Minneapolis protests, the city instituted disappointingly limited policy changes.[44] But there are more encouraging results. For example, public opinion about police has shifted dramatically, with calls to defund the police in particular increasing in popularity.[45] We should not be too quick to dismiss the possibility that rioting can achieve worthwhile political goals *qua* expressive protest.

Finally, per the proportionality constraint, political rioting can be targeted relatively narrowly: in Minneapolis, the police station, not the fire department, was torched. These sorts of distinctions are not only expressively significant but also necessary to satisfy the proportionality constraint. Under

41 Pasternak, "Political Rioting," 387. Of course, there might be situations where political rioting is, in fact, the only efficacious option. But note that Pasternak never analyzes political rioting in connection with political revolution. Perhaps this is simply outside the scope of her project, or perhaps she thinks that once political revolution breaks out and war is in the streets, we cannot be said to live in a properly democratic society.

42 Political rioters report feeling that rioting is "the only way to make [themselves] heard" (Waddington, "The Madness of the Mob?," 685).

43 Note that in general, by the necessity and success constraints, we need not necessarily exhaust less extreme methods such as lawful protest or civil disobedience before rioting if these lesser harms will not be efficacious.

44 Herndon, "How a Pledge to Dismantle the Minneapolis Police Collapsed."

45 Fleming-Wood, Margalit, and Schaffner, "Support for Cutting Law Enforcement Funding Has Spiked in the Wake of the Recent Protests."

wide proportionality, a certain amount of harm to bystanders may count as acceptable collateral damage, although contextual features will figure prominently here. For example, smashing a Chase Bank window or looting a Target is very different from burning down a local mom-and-pop restaurant owned and beloved by deeply aggrieved citizens.[46]

Admittedly, the proportionality constraint is the trickiest for political rioters to navigate on the ground. While Pasternak is right that riots are much less capable of inflicting collateral damage than armies, rioters may still endanger local shops, homes, and neighbors.[47] How can the harms caused by rioters be kept proportional to the harms they aim to avert? Armies face a similar problem. It is surely not morally sufficient to plan war crime tribunals for *after* the revolution—moral duties apply to agents during political action as well. But because political riots lack clear hierarchies of command, I think the responsibility of individual political rioters is arguably *increased* relative to that of soldiers. Thus, if one rioter begins to act outside the scope of the necessity, success, or proportionality constraints, it is principally incumbent on other rioters to intervene then and there.[48] Additionally, political rioters may incur duties to assist harmed members of their community after the riot is over.[49]

Let us try putting this analysis into action. Assume, for a moment, the position of a deeply aggrieved citizen. For generations, your community has complained to the state about how it treats you, to little material effect. You have run up against the limits of the transformative potential of complaining to *X* about *X*.

Given this context, for you to resist the apparent mechanisms of institutional justice in your society by uncivil means, for you to create a targeted zone of lawlessness where the state's ongoing war against you can be acted out in physical miniature, expresses your wholesale rejection of the political status

46 While Pasternak notes that targeting businesses may cloud the political rioter's expressive intentions, "an important exception here concerns the property of private agents who are themselves inexcusably complicit in the injustice against which the rioters protest" ("Political Rioting," 404). I agree that a similar line of argument could function against complicit corporations, potentially justifying the looting of the Minneapolis Target, although pursuing this would lie outside the scope of this project. For relevant background, see Mak, "Target Has a Long History with the Minneapolis Police."

47 Pasternak, "Political Rioting," 415.

48 Compare Pasternak's empirically grounded discussion of "crowd norms" ("Political Rioting," 414–15), or Havercroft's report that "historians of crowd behavior have long demonstrated that crowds have their own 'moral economy'" and "tend to behave well in protests" ("Why Is There No Just Riot Theory?," 913).

49 At stake here is not only the moral justification of the riot in terms of necessity and proportionality but also its expressive clarity—that is, one of the primary drivers of its success.

quo as unbearably illegitimate. It expresses the deep disrespect that has already long been mutual between the state and yourself. It expresses how serious you are that your only further recourse is that of political revolution, a complete *severing* of your protestive relationship with the state, a recourse you are not yet inclined to take. In this way, it might even express the lingering hope that your relationship with the state might be salvaged yet. It is precisely this hopefulness—this sense that your relationship with the state is still worth *fighting* for—that is overlooked when we view rioters as nothing more than reckless criminals or overeager revolutionaries.

Often, a political order that persistently refuses to hear or address the grievances of its deeply aggrieved citizens through established, ordinary channels can only be shaken to attention by unestablished, out-of-the-ordinary means.[50] Once we read political rioting as an expressive *protest* tinged with hope and optimism for a better future relationship with the state, we may be much more sympathetic to the moral case for political rioting. Importantly, we can at least see how lazy critiques of these political rioters as *lawless* miss the entire expressive point of their actions.

4. Success in Material and Existential Terms

How should we judge the success of political rioting, particularly when staring down the overwhelming force of the state? Suppose that you and I are deeply aggrieved citizens living under an unjust state in the not-too-distant future. The state—knowledgeable as it is of the contemporary defensive harm literature—decides to militarize its police to a nearly unimaginable degree. Our living conditions become increasingly impoverished as more and more resources are funneled into law enforcement, to the point where we citizens have virtually no prospect of successfully mounting public political protest of any kind. Robo-police effortlessly disperse the merest beginnings of any assembly with almost unnoticeably effective technological force; coordinated media blackouts ensure that, even if a large-scale collective protest *did* occur, its expressive reach would be extraordinarily stunted. On every front, our dystopian state ensures that public political protest is practically impossible. Would the state thereby render public protest *morally impermissible* for us as citizens, given the success constraint? Would we, the ruthlessly oppressed, be morally required to stand by?

50 Compare Hayward's "Disruption" on how political disruption can shake the politically comfortable out of their motivated ignorance to attend to serious injustices.

In this section, I hope to leverage this initial intuition pump into broadening our criteria under which political protests should count as successful. If my arguments stand, then political protest in general—and political rioting in particular—meets the success constraint in a much wider range of circumstances than we might otherwise expect. As a result, political protest cannot be morally straitjacketed by overwhelming state militarization alone.

But we need not turn to science fiction to find critics of the success constraint. Saul Smilansky notes that it gives rise to deeply paradoxical results in real life:

> In general, the more ruthless the aggressor, the more difficult it is to stop him from carrying out his threat. As a result, [the success constraint] is probably met less in ruthless aggressors than in more merciful ones. This implies that the more ruthless the aggressor, the less justified the victim would be in any attempt to kill him.[51]

Daniel Statman argues that the success constraint

> demands submission to evil and passivity in the face of wickedness. If this is what some moral or legal theory demands of us, it seems like a reductio of the theory.[52]

The intuition that defensive harm can be justified, even absent reasonable prospects of material success, has been formulated in a variety of ways. Honor-based accounts such as Statman's suggest that in the face of hopeless odds, defensive harm may be justified as an effort to uphold and defend the victims' *honor*—that is, not only their own sense of themselves but also others' sense of them as having value and not being mere objects for use. Sometimes, Statman explains, "we feel we must protect not only our body or our property but our *selves*."[53] And in these cases, violent force may be the only recourse that remains. But these accounts are both controversial and difficult to elaborate without allowing honor to take on the perverse role that it does in (falsely) justifying honor killings, as both Pasternak and Statman himself readily acknowledge.[54]

Even so, I think there is something deeply right about honor-based accounts. Statman's article is written with the paradigm case of the Warsaw ghetto uprising in mind, a calamitous yet noble effort by Jews facing Nazi extermination to "die

51 As personally related to Daniel Statman ("On the Success Condition for Legitimate Self-Defense," 666).

52 Statman, "On the Success Condition for Legitimate Self-Defense," 664.

53 Statman, "On the Success Condition for Legitimate Self-Defense," 668.

54 Pasternak, "Political Rioting," 399; and Statman, "On the Success Condition for Legitimate Self-Defense," 670.

with a gun in their hands rather than in Treblinka or another death camp."[55] In a footnote, Statman approvingly quotes Rachel L. Einwohner's assessment that the goal "was to act honorably"—not, *per impossibile*, to vanquish the Germans.[56]

Here Rodin objects, arguing that "if inflicting harm on *A* would not prevent, delay, or ameliorate the threatened harm in any way, then it is hard to see how *A* could be liable to the harm as a matter of defense."[57] Although I grant the objection, the assertion of oneself as a person with dignity and certain basic liberties is not reducible to purely material gain or loss. When *A* harms *B* by denying *B*'s very personhood—say, when *A* (a state) actively or passively maintains *B*'s status as a deeply aggrieved citizen whose suffered injustices will not be redressed in the foreseeable future—*B* is harmed in respect-based or *existential* terms.[58] And then, for *B* to "soldier on," even to die on her feet, may be morally permissible or even praiseworthy, not because she acts for the sake of her Honor (some mysterious noun in the heavens) but because doing so expressively asserts and honors—as a verb!—her own personhood in her relations with others and the state here on earth.

Instead of defending their *honor*, I submit that actors in these desperate cases are expressively reasserting their *personhood* to themselves and to one another. This distinction is worth making clear: it is the difference between honor and dignity. What is at stake is not my *honor*, the respect due to me for the kind of person I am, but my *dignity*, the respect due to me for being a person at all.[59] And in general, my dignity carries great weight not just for me but for my society as a whole. As Delmas argues, "if the law's failure to respect everyone's dignity is sufficiently threatening or destructive, all people, not just those affected by indignity, may demand reform or revolution."[60] Undermining my honor is largely a local offense, but undermining my dignity has consequences for all.

Pasternak primarily understands the success constraint in terms of whether rioters are able to influence the *policies* underlying "material deprivation and social exclusion."[61] But I suggest there is another way in which rioters can be successful, even beyond resisting marginalization or communicating anger and defiance, which Pasternak considers but does not champion:

55 Statman, "On the Success Condition for Legitimate Self-Defense," 665.

56 Einwohner, "Opportunity, Honor, and Action in the Warsaw Ghetto Uprising of 1943," 666, cited in Statman, "On the Success Condition for Legitimate Self-Defense," 665.

57 Rodin, "Justifying Harm," 93.

58 I hope to bracket concerns relating to the group agency of *A*. At most, they should affect the form and not the content of this analysis.

59 Here I bracket whether nonpersons can have dignity.

60 Delmas, *A Duty to Resist*, 178.

61 Pasternak, "Political Rioting," 398.

> Perhaps, it can be argued that the success condition is fulfilled even if rioters do not have a reasonable prospect of achieving all their goals. Perhaps, it would be enough, for example, if they have a reasonable prospect of resisting political marginalization and communicating anger and defiance, thus maintaining a sense of self-respect and pride. Some accounts of permissible defensive harm would support this conclusion, as they suggest that victims of aggression can be justified in inflicting harm on their aggressors even when doing so would have no chance of mitigating the original attack, if through their actions they demonstrate that they are not "just passive objects to be trodden upon." But this position strikes many as controversial and anyway will be even less persuasive if the rioters would in fact worsen the condition of fellow oppressed citizens.[62]

Here is a defense of this controversial position. Even when tactical defeat in material terms really is inevitable and the broader expressive reach of a political riot will be quashed, the higher-level goal of existential self-assertion always seems available and valuable for its own sake. By bringing attention to indignities suffered by the deeply aggrieved, political rioters uphold the conditions of the state's legitimacy better than the state does itself, challenging the state to do better. This is the optimistic thrust of the political riot: unlike revolutionaries, rioters implicitly reaffirm that greater justice and legitimacy are achievable for *this* state. They have not (yet) abandoned the state's political project. In this way, political rioting can even be *healthy* for an unjust state.

There may still be times when inaction is morally required by the proportionality constraint, if other harms incurred by action would be bad enough. (For an extreme example, suppose the state threatened not just to suppress our protest but to slaughter our entire neighborhood if any one of us spoke out.) And the necessity constraint may require us to pursue lesser methods of protest if they too could reasonably attain success in material or existential terms. But now, these are questions to be weighed and considered, not assumed improper in advance.

To highlight this, I turn to the well-known case of Judy Norman, who was physically and mentally abused by her husband for twenty years.[63] He regularly made her prostitute herself, starved her, and broke glass against her face, among infinitely many other despicable evils. As his death threats toward her became more direct, public, and unmistakable, Judy Norman thoroughly exhausted all legal avenues available to her to try to save her own life. She repeatedly tried to escape, called the police until they no longer came, and attempted to have

62 Pasternak, "Political Rioting," 399.

63 My summary draws from State v. Norman 324 N.C. 253, 378 S.E.2d 8 (1989).

her husband committed to a mental health center before fatally shooting him in his sleep.

It is generally accepted that in the Norman case, and cases like it, the victim is fully justified in using violent or even lethal force to defend herself.[64] And our defensive harm analysis delivers this result. First, it seems clear that Norman's actions were necessary in both material terms (to save her life) and existential terms (to reassert her personhood); she had exhausted every other avenue available to her. Second, her prospects of success were very good, not only in material terms (she was incredibly likely to succeed at killing her sleeping husband) but in existential terms as well (she was guaranteed to reassert her own personhood just by continuing to fight for her own survival). Even if the past material and existential harms she suffered could not be undone and would continue to impact her, she could still prevent further harms to herself. Finally, her use of force against her husband was clearly proportionate in light of his increasingly serious death threats, even before considering the rest of his material and existential abuse. Her husband was no bystander; he was fully liable under narrow proportionality.

In cases of individual self-defense, existential prospects of success spring to the fore of our considerations. But cases of *collective* self-defense are often much more complicated. In particular, note that the proportionality constraint, which calls on us to distinguish between bystanders and liable parties as best as we are able, may be more challenging to apply in the case of a political riot, where civilian bystanders might find themselves caught up in the violence and liability for structural injustices may be difficult to assign to particular individuals. This suggests that political rioters should target state property and clearly liable state agents as narrowly as possible.[65]

But the analogy to the Norman case highlights that deeply aggrieved citizens can legitimately claim that they do not bring violence to the table *ex nihilo*.[66] At its best, political rioting expresses that the maintenance of deeply aggrieved cit-

64 See, for instance, Helen Frowe's *Defensive Killing*, 140–41, or Jeff McMahan's "War as Self-Defense," 76, on cases with this structure.

65 Given the ends of paradigmatic political rioting (seeking greater union with the state), I follow Pasternak in thinking that *killing* police officers would almost always be deeply expressively counterproductive ("Political Rioting," 405). Note too that the relationship between deeply aggrieved citizens and police is at the very least more mediated than the relationship between Judy Norman and her husband.

66 Note just how much is at stake when determining where this violence originates. State actors may point to the spontaneous public violence of rioters as reason to overwhelm them with force. But I have argued that justified political rioters may be using material violence to ocularize ongoing existential violence—and indeed, even to protest ongoing material violence at the hands of the riot police responding to them. Once we read justified

izens in their positions of relative subordination is itself the principal material and existential harm that is to be averted. And this is a level of standing violence at the hands of the state that thoroughly permeates the lives of deeply aggrieved citizens. This is because the *threat* of violence is itself already violence (if you doubt this, you have never been mugged). And to be a deeply aggrieved citizen is to live under the standing threat that your basic liberties may be violated, without proper restitution. The graveness of this existential harm is such that even significant material defensive harms may be proportionate in response.

Indeed, the existential side of the ledger is actually *more* fundamental than the material. We have already seen Pasternak argue that political rioters can affect public policy, resist marginalization, and express angry defiance.[67] D'Arcy stresses that militant protest may be required to uphold and even reclaim the democratic ideal of the people's self-governance.[68] And Havercroft emphasizes that political rioting can extra-institutionally preserve freedom, promote equality, and give voice to the grievances of marginalized groups.[69] But what underlies all three of these analyses is a firm commitment to the existential import of respecting the dignity of persons—this is core not only to our notion of justice itself but also to the state's own claims to legitimacy. A complete moral accounting of the status of riots should give a central place to the existential benefits and harms that unite and underlie all these considerations.

Violence comes with great costs, even when put to worthwhile political ends with the best of intentions.[70] Beyond the direct harms of violence itself, violent tactics may limit the public's ability or willingness to support or join protestors

political riots as *defensive* harms necessitated by ongoing existential violence, we can see that in a deeper sense, responsibility rests on those perpetuating these injustices.

67 Pasternak, "Political Rioting," 387.

68 D'Arcy, *Languages of the Unheard*, 72.

69 Havercroft, "Why Is There No Just Riot Theory?," 913.

70 This is why *right*-wing agitators were caught vandalizing local businesses and firing bullets into the Third Precinct (Beckett, "'Boogaloo Boi' Charged in Fire of Minneapolis Police Precinct during George Floyd Protest"; Peiser, "'Umbrella Man' Went Viral Breaking Windows at a Protest"). These right-wing agitators intended to incite violence that was truly disorderly (or "random"), would confuse the political expression of the protestors, and would invite the violence of riot police upon them. There are two distinct kinds of harms here. Materially, these agitators not only smashed windows and lit fires but invited the escalation of riot police against their political enemies. But existentially, they also knew that the appearance of open violence in the streets would be used to fuel perceptions of political rioters as disorderly "criminal types" who were only capable of expressing themselves in this way, further disrespecting them as persons and directing public attention away from the expressive nature of their protests.

and may even appear to legitimize violent state repression in response.[71] These are important costs to weigh when determining the proportionality of political riots. But while the existing literature has focused on these sorts of costs and various material benefits, our moral accounting should fully acknowledge *all* relevant benefits and harms—and in particular, the central existential benefit of reasserting one's own dignity as a person and the tremendous existential harm of continuing to languish in disrespect.

Dara T. Mathis reminds us that "when violent state actors preemptively call for nonviolence to manipulate protestors to comply without addressing their grievance, nonviolence is another way to muzzle the voiceless."[72] In these cases, calls for civility inappropriately silence legitimate and urgent public expression. A political riot may provide a necessary, successful, and proportionate public forum for deeply aggrieved citizens to ocularize their warranted disrespect for the state that maintains them in ongoing subjection, as well as their inviolable respect for themselves as persons with dignity *beyond* the boundaries of civility. We must remain wary of arguments against political rioting that overlook the significance of systemic material and existential harms in favor of upholding civility at any cost, thereby preferring "a negative peace which is the absence of tension to a positive peace which is the presence of justice."[73]

Indiana University
rimouser@indiana.edu

REFERENCES

Barker, Kim, Mike Baker, and Ali Watkins. "In City after City, Police Mishandled Black Lives Matter Protests." *New York Times*, March 20, 2021. https://www.nytimes.com/2021/03/20/us/protests-policing-george-floyd.html.

Beckett, Lois. "'Boogaloo Boi' Charged in Fire of Minneapolis Police

71 Although they focus primarily on movements with politically revolutionary goals, see Chenoweth and Stephan's *Why Civil Resistance Works* for empirical argument that these concerns should incline us toward adopting nonviolence.

72 Mathis, "King's Message of Nonviolence Has Been Distorted."

73 King, "Letter from a Birmingham Jail." This paper has benefited tremendously from comments and discussion with Matthew Adams, Zara Anwarzai, Marcia Baron, Gary Ebbs, Kjell Fostervold, Paul Howatt, Savannah Pearlman, John Robison, Paul Shephard, Kyle Stroh, Elizabeth Williams, and anonymous reviewers.

Precinct during George Floyd Protest." *Guardian*, October 23, 2020. https://www.theguardian.com/world/2020/oct/23/texas-boogaloo-boi-minneapolis-police-building-george-floyd.

Caputo, Angela, Will Craft, and Curtis Gilbert. "'The Precinct Is on Fire': What Happened at Minneapolis' 3rd Precinct—and What It Means." *MPR News*, June 30, 2020. https://www.mprnews.org/story/2020/06/30/the-precinct-is-on-fire-what-happened-at-minneapolis-3rd-precinct-and-what-it-means.

Celikates, Robin. "Democratizing Civil Disobedience." *Philosophy and Social Criticism* 42, no. 10 (December 2016): 982–94.

Chan, David K. *Beyond Just War*. Basingstoke: Palgrave Macmillan, 2012.

Chenoweth, Erica, and Maria J. Stephan. *Why Civil Resistance Works: The Strategic Logic of Nonviolent Conflict*. New York: Columbia University Press, 2011.

Cherry, Myisha "Value-Based Protest Slogans: An Argument for Reorientation." In T*he Movement for Black Lives: Philosophical Perspectives*, edited by Michael Cholbi, Brandon Hogan, Alex Madva, and Benjamin Yost, 160–75. New York: Oxford University Press, 2021.

Cobb, Charles E., Jr. *This Nonviolent Stuff'll Get You Killed*. New York: Basic Books, 2014.

D'Arcy, Stephen. *Languages of the Unheard: Why Militant Protest Is Good for Democracy*. London/New York: Zed Books, 2013.

Delmas, Candice. *A Duty to Resist: When Disobedience Should Be Uncivil*. New York: Oxford University Press, 2018.

Einwohner, Rachel L. "Opportunity, Honor, and Action in the Warsaw Ghetto Uprising of 1943." *American Journal of Sociology* 109, no. 3 (November 2003): 650–75.

Epstein, Reid J. "Full Transcript: Mike Pence's R.N.C. Speech." *New York Times*, August 26, 2020. https://www.nytimes.com/2020/08/26/us/politics/mike-pence-rnc-speech.html.

Fleming-Wood, Bennett, Yonatan Margalit, and Brian Schaffner. "Support for Cutting Law Enforcement Funding Has Spiked in the Wake of the Recent Protests." *Data for Progress*, July 7, 2020. https://www.dataforprogress.org/blog/2020/7/3/support-for-cutting-law-enforcement-funding-has-spiked.

Frowe, Helen. *Defensive Killing: An Essay on War and Self-Defence*. Oxford: Oxford University Press, 2014.

Havercroft, Jonathan. "Why Is There No Just Riot Theory?" *British Journal of Political Science* 51, no. 3 (July 2021): 909–23.

Hayward, Clarissa Rile. "Disruption: What Is It Good For?" *The Journal of Politics* 82, no. 2 (April 2020): 448–59.

Herndon, Astead W. "How a Pledge to Dismantle the Minneapolis Police Collapsed." *New York Times*, September 26, 2020. https://www.nytimes.

com/2020/09/26/us/politics/minneapolis-defund-police.html.
Hill, Thomas E., Jr. "Symbolic Protest and Calculated Silence." *Philosophy and Public Affairs* 9, no. 1 (Autumn 1979): 83–102.
Hurka, Thomas. "Proportionality in the Morality of War." *Philosophy and Public Affairs* 33, no. 1 (Winter 2005): 34–66.
Kaul, Greta. "Seven Days in Minneapolis: A Timeline of What We Know about the Death of George Floyd and Its Aftermath." *MinnPost*, May 29, 2020. https://www.minnpost.com/metro/2020/05/what-we-know-about-the-events-surrounding-george-floyds-death-and-its-aftermath-a-timeline/.
King, Martin Luther, Jr. "Letter from a Birmingham Jail." Center for Africana Studies, University of Pennsylvania, 1963. https://www.africa.upenn.edu/Articles_Gen/Letter_Birmingham.html.
Lai, Ten-Herng. "Justifying Uncivil Disobedience." In *Oxford Studies in Political Philosophy*, vol. 5, edited by David Sobel, Peter Vallentyne, and Steven Wall, 90–114. Oxford: Oxford University Press, 2019.
Lazar, Seth. "Necessity in Self-Defense and War." *Philosophy and Public Affairs* 40, no. 1 (Winter 2012): 3–44.
Lyons, David. "Moral Judgment, Historical Reality, and Civil Disobedience." *Philosophy and Public Affairs* 27, no. 1 (Winter 1998): 31–49.
Mak, Aaron. "Target Has a Long History with the Minneapolis Police." *Slate*, May 29, 2020. https://www.slate.com/business/2020/05/targets-long-history-with-minneapolis-police.html.
Mathis, Dara T. "King's Message of Nonviolence Has Been Distorted." *The Atlantic*, April 3, 2018. https://www.theatlantic.com/politics/archive/2018/04/kings-message-of-nonviolence-has-been-distorted/557021/.
McMahan, Jeff. *Killing in War*. Oxford: Oxford University Press, 2009.
———. "War as Self-Defense." *Ethics and International Affairs* 18, no. 1 (March 2004): 75–80.
Moraro, Piero. "On (Not) Accepting the Punishment for Civil Disobedience." *Philosophical Quarterly* 68, no. 272 (July 2018): 503–20.
Nodjimbadem, Katie. "The Long, Painful History of Police Brutality in the U.S." *Smithsonian Magazine*, July 27, 2017. Updated May 29, 2020. https://www.smithsonianmag.com/smithsonian-institution/long-painful-history-police-brutality-in-the-us-180964098/.
Pasternak, Avia. "Political Rioting: A Moral Assessment." *Philosophy and Public Affairs* 46, no. 4 (Fall 2018): 384–418.
Peiser, Jaclyn. "'Umbrella Man' Went Viral Breaking Windows at a Protest. He Was a White Supremacist Trying to Spark Violence, Police Say." *Washington Post*, July 29, 2020. https://www.washingtonpost.com/nation/2020/07/29/umbrella-man-white-supremacist-minneapolis/.

Pineda, Erin. "Civil Disobedience and Punishment: (Mis)reading Justification and Strategy from SNCC to Snowden." *History of the Present* 5, no. 1 (Spring 2015): 1–30.

Rawls, John. *A Theory of Justice*, Cambridge, MA: Belknap Press, 1971.

Rodin, Daniel. "Justifying Harm." *Ethics* 122, no. 1 (October 2011): 74–110.

Rothman, Lily. "What Martin Luther King Jr. Really Thought about Riots." *Time*, April 28, 2015. https://www.time.com/3838515/baltimore-riots-language-unheard-quote.

Scheuerman, William E. *Civil Disobedience*. Cambridge: Polity Press, 2018.

———. "Why Not Uncivil Disobedience?" *Critical Review of International Social and Political* Philosophy (November 2019): 1–20.

Schwartz, Stephan. A. "Police Brutality and Racism in America." *Explore* 16, no. 5 (September–October 2020): 280–82.

Statman, Daniel. "On the Success Condition for Legitimate Self-Defense." *Ethics* 118, no. 4 (July 2008): 659–86.

Tavernise Sabrina, and Matthew Rosenberg. "These Are the Rioters Who Stormed the Nation's Capitol." *New York Times*, January 7, 2021. https://www.nytimes.com/2021/01/07/us/names-of-rioters-capitol.html.

Taylor, Derrick Bryson. "George Floyd Protests: A Timeline." *New York Times*, 2021. https://www.nytimes.com/article/george-floyd-protests-timeline.html.

Waddington, David. "The Madness of the Mob? Explaining the 'Irrationality' and Destructiveness of Crowd Violence." *Sociology Compass* 2, no. 2 (March 2008): 675–87.

Journal of Ethics and Social Philosophy
VOL. 26, NO. 3 · FEBRUARY 2024
https://doi.org/10.26556/jesp.v26i3.2844

DISMISSING BLAME

Justin Snedegar

When someone blames you, there are various ways you might respond. First, you might *accept* blame. You agree that you are blameworthy, which means you agree that you have done something wrong and that you do not have an adequate excuse or exemption.[1] You will typically feel emotions such as guilt or remorse and take reparative steps by apologizing or making amends.[2] Second, you might *reject* blame by denying that you are blameworthy. If you reject blame, you are unlikely to feel guilty or take reparative steps, since you think no moral repair is necessary. Both of these are direct responses to being blamed: both involve *engaging* with the blame and, in particular, with the blamer by either agreeing that you are blameworthy and reacting appropriately or else explaining why you are not blameworthy.[3] When we engage with blame directly, either by accepting or rejecting it, there can be

1 We may accept blame in one sense without thinking that we have acted wrongly. This sense is most familiar when some bad outcome is the result of a group's actions and someone steps up to take the blame, even if they are not plausibly responsible for the bad outcome; see, e.g., Collins, "Filling Collective Duty Gaps." As Stephen Bero observes, we may also accept blame in some sense in the individual case when someone *mistakenly* believes that we have done something wrong. Accepting blame can be a shortcut to smoothing things over. As Bero also observes, the pressure to do this will be distributed according to unfortunately familiar social hierarchies ("Holding Responsible and Taking Responsibility," 291–92).

2 As we will see, agreeing that you are blameworthy, feeling guilty, and taking reparative steps may often be necessary but are not sufficient for accepting blame. There may be cases in which there is nothing in particular you should do in response to your wrongdoing other than acknowledging it and trying to do better going forward. For example, in some cases, it will be too late for any meaningful moral repair.

3 For the language of direct and indirect responses, see Cohen, "Casting the First Stone," 119; and Lippert-Rasmussen, *Relational Egalitarianism*, 96. James Edwards instead talks about *content-sensitive* vs. *content-insensitive* responses ("Standing to Hold Responsible," 448). For discussion of the menu of responses to blame, see Walker, *Moral Repair*, 135; McKenna, *Conversation and Responsibility*, 88–89; Bell, "The Standing to Blame," 264; and Friedman, "How to Blame People Responsibly," 275. Daniela Dover is critical of the call-and-response model of blaming or critical interactions implicit in some of this discussion ("Criticism as Conversation").

positive moral upshots, including opportunities for moral repair, taking a stand for our values, and engaging in edifying moral discussion with others.

This paper is about a third way you might respond to being blamed, which I call *dismissing* blame; alternatively, we could call it "brushing off" or "disregarding" blame. In contrast to both accepting and rejecting blame, dismissing blame is an *indirect* response, because it does not involve engaging with the blamer about the content of the blame; it is a *refusal to engage* with the blamer about the (supposed) wrongdoing. Many think that this response is appropriate when the person blaming you is doing so hypocritically or when it is none of their business. At least, it is the response that people often give in such circumstances: "Who are *you*, of all people, to blame me for this?" This does not mean that you do not believe that you are blameworthy. You might preface the dismissal by admitting that you have acted wrongly: "Sure, I shouldn't have done it. But who are *you* to blame me for it?" You might be perfectly willing to accept blame from *other* people and to undertake moral repair. But if you dismiss blame from someone, then you will not engage with *their* blame. This paper defends an account of what it is to dismiss blame.

The phenomenon is quite widespread and can come in many forms, just as blame itself can come in many forms. If someone blames you verbally and face to face, the most obvious way we can dismiss blame is by responding to the blamer with something like "Who are *you* of all people to blame me?" Other kinds of dismissal will be appropriate given other kinds of blame. For example, if you shoot me a nasty look because you think I cut you off in traffic, rather than signaling to you that I accept blame by giving you a sheepish wave, I might instead roll my eyes and wave my hand in a dismissive way, especially if I only cut you off because you had cut me off a moment ago. We can also dismiss blame that is not expressed to us. If a third party informs me that you—a habitual liar—have been blaming me for some recent dishonesty, I might express dismissal of this blame to that third party, saying "That hypocrite! Who are they to blame me for this?" I might even find out, for example, by reading your diary, that you have been blaming me for something for which you lack standing. Just as you kept your blame to yourself, I can keep my dismissal of that blame to myself.[4] My focus is on face-to-face, direct blaming interactions, but I return to these nondirect cases at the end of section 4.[5]

4 Thanks to an anonymous referee for pressing me to say more about the scope of the phenomenon and for providing some of these nice examples.

5 Many authors (though not all—see, e.g., Herstein, "Justifying Standing to Give Reasons," n12) think that one can lack standing even for *private* blame—blame kept to oneself—though almost all focus on expressed, direct blame; see, e.g., Wallace, "Hypocrisy, Moral Address, and the Equal Standing of Persons"; Todd, "A Unified Account of the Standing

Here are two reasons why moral philosophers might be interested in this topic. First, there has been much recent work on the ethics of blame—questions about when it is appropriate for some particular person or group to blame another particular person or group, even granting that the latter is blameworthy. Much of this work has focused on when and why the blamer has or lacks the *standing* to blame.[6] An account of dismissing blame should inform these debates, because the standing to blame is often characterized in terms of dismissing blame: what is distinctive of standingless blame is that you can legitimately dismiss it.[7] In contrast, as Macalester Bell points out, other ways that blame can be inappropriate, such as being overly harsh, badly timed, or petty, do not seem to license dismissing blame but only objecting to the tone or the timing.[8] My focus is on dismissing blame, but at the end of the paper I briefly explore how the account here may bear on important questions about the standing to blame.

An account of what it is to dismiss blame will also bear on equally important but less well-studied questions about the ethics of *responding* to blame. There are questions about when and why it is legitimate to dismiss blame. Sometimes it seems that we are within our rights to dismiss blame from someone, but often we are not, and many actual cases in which blame is dismissed fall into the latter

---

to Blame"; Fritz and Miller, "Hypocrisy and the Standing to Blame." Other authors who do not take an explicit stand on this question nevertheless focus on expressed blame, e.g., Bell, "The Standing to Blame"; Dover, "The Walk and the Talk"; and Edwards, "Standing to Hold Responsible," 441.

6 For a sampling of recent work on the standing to blame, see Cohen, "Casting the First Stone" and "Ways of Silencing Critics"; Smith, "On Being Responsible and Holding Responsible"; Duff, "Blame, Moral Standing, and the Legitimacy of the Criminal Trial"; Wallace, "Hypocrisy, Moral Address, and the Equal Standing of Persons"; Friedman, "How to Blame People Responsibly"; Bell, "The Standing to Blame"; Herstein, "Understanding Standing" and "Justifying Standing to Give Reasons"; Isserow and Klein, "Hypocrisy and Moral Authority"; Fritz and Miller, "Hypocrisy and the Standing to Blame" and "The Unique Badness of Hypocritical Blame"; Roadevin, "Hypocritical Blame, Fairness, and Standing"; Rossi, "The Commitment Account of Hypocrisy"; Todd, "A Unified Account of the Standing to Blame" and "Let's See You Do Better"; Dover, "The Walk and the Talk"; Edwards, "Standing to Hold Responsible"; Piovarchy, "Hypocrisy, Standing to Blame and Second-Personal Authority"; Tognazzini, "On Losing One's Moral Voice"; King, "Skepticism about the Standing to Blame"; Rivera-López, "The Fragility of Our Moral Standing to Blame"; and many others.

7 On the "deflection test" for standing, see especially Edlich, "What about the Victim?," 213. See also, e.g., Linda Radzik's talk of "dismissing" both the blamer and the content of the blame in "On the Virtue of Minding Our Own Business," 178; Edwards on "dismissing" accusations in "Standing to Hold Responsible," 447; G. A. Cohen's talk of "silencing" critics in "Casting the First Stone" and "Ways of Silencing Critics"; Marilyn Friedman's talk of "ignoring" blame in "How to Blame People Responsibly," 282; and Ori Herstein's talk of "disregarding" blame in "Understanding Standing" and "Justifying Standing to Give Reasons."

8 Bell, "The Standing to Blame."

category.[9] Consider cases in which someone issues a charge of hypocrisy or tells someone to mind their own business as a diversionary tactic to escape criticism. There are also important questions about the wrongs we commit when we illegitimately dismiss blame and how these wrongs interact with other sorts of injustice. For example, Sue Campbell discusses the illegitimate dismissal of women's moral complaints on the basis of "bitterness" or "emotionality."[10] To address these ethical questions, we need a clear understanding of what it is to dismiss blame. Though I touch on some of these ethical issues at various points, my focus is on the conceptual question of what it is to dismiss blame rather than on when doing so is appropriate.

In the next section, I describe a useful starting point for theorizing about dismissing blame. This is the popular idea that blaming involves making demands of the blamed party. According to this view, dismissing blame involves dismissing demands issued by blame. I then consider various accounts of exactly which demands we dismiss when we dismiss blame. Many authors have mentioned potential answers to this question in passing, often in discussions of the standing to blame or of the nature of blame itself. I argue that all of them face problems or at least leave important questions unanswered. I use lessons from the discussion of these views to develop my own proposal: to dismiss blame is to dismiss a demand for a second-personal expression of remorse *to* the blamer.

1. DISMISSING BLAME AS DISMISSING DEMANDS

I assume that blaming someone involves, among other things, issuing implicit demands to that person. This is certainly not uncontroversial, but it is a widely accepted way of thinking about blame and one that philosophers with otherwise importantly different views of blame can accept and have defended.[11] This

9 Authors who are skeptical about losing the standing to blame may likewise be skeptical about whether dismissing blame, in the sense at issue in this paper, is legitimate. See, e.g., Bell, "The Standing to Blame"; Dover, "The Walk and the Talk"; and King, "Skepticism about the Standing to Blame." On the abuse of dismissing blame via a charge of hypocrisy in political contexts, see McDonough, "The Abuse of the Hypocrisy Charge in Politics"; and O'Brien and Whelan, "Hypocrisy in Politics." Herstein observes that the practice of invoking standing to dismiss blame (and other interventions) is "precarious" because it is tempting to use it illegitimately ("Justifying Standing to Give Reasons," 18).

10 Campbell, "Being Dismissed." See also Carbonell, "Social Constraints on Moral Address."

11 For philosophers who accept some version of this idea, see Strawson, "Freedom and Resentment"; Watson, "Responsibility and the Limits of Evil"; Wallace, *Responsibility and the Moral Sentiments* and "Emotions, Expectations, and Responsibility"; Hieronymi, "The Force and Fairness of Blame"; Walker, *Moral Repair*; Darwall, *The Second-Person Standpoint*; Smith, "Control, Responsibility, and Moral Assessment"; Duff, "Blame,

way of thinking about blame is an assumption of this paper, but one relevant attraction is that it gives us a promising way of explaining the sense in which blame goes beyond simply grading an agent's actions or pointing out wrongdoing to them.[12] In the context of this paper, it helps us see why agents might be eager to dismiss blame. The ability to dismiss mere grading or pointing out of wrongdoing does not seem to capture the appeal of dismissing blame.

There are different ways of developing this picture, but the general idea gives us a natural way to think about the standing to blame, since issuing demands is something that we can have or lack the standing to do.[13] A higher-ranking military officer has the standing to issue demands to a lower-ranking officer but not vice versa. A parent has the standing to issue certain demands to her child but not vice versa. These examples illustrate the standing to issue a demand but are arguably not so helpful for thinking about the standing to blame, since they centrally involve hierarchical relationships. These kinds of relationships do not hold between mature moral agents outside of special relationships such as parent-child or commanding officer–subordinate, and clearly, we can blame one another outside of these kinds of relationships. But as Darwall and others emphasize, it is plausible to think about morality—at least a large part of it—as involving second-personal demands between peers.[14] If so, then we can think of blame as involving demands that we make on one another even when there are no hierarchical relationships involved.

Moral Standing, and the Legitimacy of the Criminal Trial"; Fricker, "What's the Point of Blame?"; Edwards, "Standing to Hold Responsible"; Piovarchy, "Hypocrisy, Standing to Blame and Second-Personal Authority"; and many others. For critical discussion, see Coleen Macnamara's "Taking Demands Out of Blame" and "'Screw You!' and 'Thank You!'" However, Macnamara can, I believe, be on board with the parts of this picture that are crucial for my purposes, since she does accept that blame calls for certain kinds of responses. Her objections are largely directed at the weightiness of demands, on the usual understanding of the term. Prominent views of blame that cannot easily accept what I say here include the views defended by George Sher in *In Praise of Blame* and T. M. Scanlon in *Moral Dimensions*. For the purposes of this paper I have to set such views aside.

12 See, e.g., Wolf, *Freedom Within Reason*, 40; and Hieronymi, "The Force and Fairness of Blame."

13 Standing is often thought of as a *right* to blame (e.g., Fritz and Miller, "Hypocrisy and the Standing to Blame"). Matt King, in "Skepticism about the Standing to Blame," argues that there is no good understanding of standing in terms of rights. Interestingly, King does not (explicitly) object to the idea that we may have a *claim right* against others that they *respond* to our blame in certain ways and that this claim right can be defeated in some cases, which would presumably license dismissing blame. His complaint is just that this does not explain why blaming without standing would be inappropriate, since the absence of claim rights does not entail impermissibility.

14 Darwall, *The Second-Person Standpoint*.

The view that blaming involves issuing demands also gives us a corresponding way to think about dismissing blame: what we dismiss when we dismiss blame are demands that the blamer has issued. Typically, a legitimate demand on us puts us under obligations or at least gives us reasons. When someone with the standing to do so issues a demand, it is something that we should pay attention to and take into account in our deliberations.[15] According to this view, we can brush off these obligations or reasons when the person blaming us lacks standing. This explains why being able to dismiss blame is a *benefit* for the blamed party. Engaging with someone about your (supposed) wrongdoing is usually unpleasant, so it matters to the blamed party that they are able to dismiss blame and so dismiss a demand that they engage in the relevant way. This explains why people tend to overuse charges of hypocrisy or meddling in an effort to dismiss blame and so escape these burdens.

It is worth pausing over the distinction between dismissing blame and *rejecting* blame. According to the view I am developing, dismissing blame is dismissing a certain demand involved in blaming. Rejecting blame, on the other hand, is denying that you are blameworthy. But whether you reject blame or dismiss it, typically you will not comply with demands issued by that blame. I will argue that the demand we dismiss when we dismiss blame is a demand for an expression of remorse to the blamer. If you dismiss this demand, then you will not comply with it. But dismissing blame cannot simply consist in not complying with this demand, since you will also not comply with it if you *reject* blame rather than dismiss it.[16]

We can understand the difference between dismissing and rejecting blame in terms of denying different preconditions or presuppositions of blame. When you reject blame, you deny that you are blameworthy. Being blameworthy is a precondition of the appropriateness of many of the demands plausibly issued by blame, for example, demands to apologize, to feel remorse, or, in my view, to express remorse to the blamer. If you explain to the blamer that you are not

15 Macnamara (in "Taking Demands Out of Blame" and "'Screw You!' and 'Thank You!'") argues that blame does not involve demands because blame can be appropriate even in cases in which issuing a demand would not be appropriate, for example, cases of "supererogation" from Julia Driver's "The Suberogatory," in which the agent acts badly but not wrongly. I do not need to assume a heavy-handed notion of demands, according to which failing to do what someone (legitimately) demands of you is necessarily impermissible. What is important is the structure: when someone legitimately demands something of you, it puts *normative pressure* on you to comply, plausibly in the form of *pro tanto* reasons. Sometimes the right thing to do, all things considered, may be to resist this pressure and *not* do what is demanded of you, e.g., if these reasons are outweighed.

16 Thanks to anonymous referees for pressing me to say more about the difference between rejecting and dismissing blame.

blameworthy, then (if they are reasonable) they will withdraw their blame as mistaken. Conversely, if you *dismiss* blame, you do not necessarily deny that you are blameworthy and so do not necessarily reject *this* precondition for blame. Rather, you deny the precondition or presupposition that the person blaming you has the authority or standing to issue the relevant demand(s).

For an analogy, imagine that you are a cleaner at a grocery store. Suppose the store manager mistakenly thinks that you are a shelf stocker and so demands that you stock the shelves. You can reject this demand, because one precondition of its appropriateness is that stocking the shelves is your job.[17] You will explain to the manager that you are a cleaner, and (if they are reasonable) they will withdraw the demand as mistaken. On the other hand, if a cashier demands that you clean the floors, you can dismiss this demand, not because cleaning the floors is not your job, and not because the floors do not need to be cleaned, but instead because the cashier is not your boss and so does not have the authority to issue this demand. In both the blame case and the cleaner case, you can, of course, both reject and dismiss the demands: you can think both that the preconditions for the relevant demands are not met and that the person issuing the demand lacks standing to do so.

Assuming that we can defend the approach of thinking about dismissing blame in terms of dismissing certain demands that blame makes on us, we still must say what those demands are. This question has received surprisingly little attention in the standing literature, with Herstein and Edwards being the main exceptions.[18] In the next section, I examine different answers to this question and argue that none are satisfactory. But the discussion brings out important lessons that inform a better account.

2. WHAT IS DISMISSED?

The question at issue is what we dismiss when we dismiss blame. In this section, I explore the idea that dismissing blame is dismissing at least one of the demands issued by blame.[19] I consider different accounts of what this demand

17 In some workplaces, it might be that if the manager demands that you do some task, even if it is technically someone else's job, you still must do it. But we can assume this is not the case at this grocery store: people have clearly defined roles, and it is in their contract that they do not have to do things outside of those roles.

18 Herstein, "Understanding Standing" and "Justifying Standing to Give Reasons"; Edwards, "Standing to Hold Responsible."

19 Some of the philosophers I draw on in this section have taken as their task giving an account of the demands issued by blame, rather than an account specifically of which demands we dismiss when we dismiss blame. It is possible, of course, that dismissing

is and argue that these fail in instructive ways. The discussion brings out three lessons that pave the way for a better account. First is the familiar point that we need to distinguish demands issued by the blamer from ordinary duties, reasons, and norms, violation of which may be the basis for blame, or which the blame may point out to us, but which hold independently of the blame. Second, what we dismiss must be a demand to do something that we have a duty or reason to do only once and because we have been blamed. Third, what we dismiss must be a demand for a second-personal response *to* the blamer; the blamer's role as *recipient* of the response is crucial. In the next section, I use these lessons to develop a promising account of dismissing blame.

2.1. Demands and Independent Moral Norms

Philosophers who think that blame involves demands have offered a range of answers to the question of what blaming someone demands of them. Some think that blame expresses a demand that the blamed party comply with moral norms. For example, Darwall says "if you express resentment to someone for not moving his foot from on top of yours, you implicitly demand that he do so." Others think that blame involves demands for reparations, such as apology or compensation, to those you have wronged. Walker holds that when we blame someone, "we demand some rectifying response" from the wrongdoer. Others, such as Shoemaker and Fricker, think that blame demands that the wrongdoer experience negative emotions constitutive of self-blame, such as remorse or guilt. Blame may demand combinations of these, as well.[20]

A striking thing about these proposals is that the things demanded are things that moral norms direct us to do independently of being blamed. Ordinary moral norms, for example, not to steal, as well as the norms directing us to apologize and to have the appropriate attitudes when we act wrongly, apply to us independently of anyone blaming us. It is widely recognized that we cannot dismiss these independent moral norms, even if the blamer lacks standing. Some real-life cases that involve dismissing blame, for example, criticizing climate activists who fly to speaking engagements on the basis of hypocrisy, are objectionable at least in part because they seem to be illegitimate attempts to

blame, in the relevant sense, only involves dismissing *some* of the demands issued by blame. So an argument that some demand is not what we dismiss does not necessarily constitute an argument that it is not issued by blame.

20 Darwall, *The Second-Person Standpoint*, 76; Walker, *Moral Repair*, 26; Shoemaker, "Moral Address, Moral Responsibility, and the Boundaries of the Moral Community," 91; and Fricker, "What's the Point of Blame?," 173.

dismiss or evade these independent moral norms.[21] So if this is the right place to look for what we can dismiss when we can dismiss blame, it is important to keep in mind that it is only the *blamer's demand* that we do these things that can be dismissed.

2.2. Blame-Specific Responses

Herstein offers an account that brings this out explicitly.[22] In Herstein's view, blaming and demanding are speech acts called *directives* and involve issuing what he calls *directive reasons*. These are reasons to do or feel certain things *because of the directive*; for the directive reason to be satisfied, the motivating reason, or the agent's basis for doing or feeling the relevant thing, must be the directive itself. When a commanding officer commands that a subordinate drop and give her twenty, this gives the subordinate a reason to drop and give her twenty and to do so because she has been commanded to do so.

Blame, in Herstein's view, gives directive reasons for the blamed party to comply with the moral norms, make reparations, or feel remorse *because of the blame*. For example, Herstein says "when Caligula blames Nero for being a bad emperor he … aims to actively give Nero reason to change his ways … as well as a reason to feel remorse, shame, responsibility and purpose (to improve), which are fitting emotional reactions to blaming." Accepting blame amounts to taking these directive reasons on board in one's deliberations about how to think, feel, and act. When the blamer lacks standing, the blamed party is permitted to dismiss the directive reasons, but not the ordinary reasons, to do these things.[23] If I hypocritically blame you for stealing, you only get to dismiss the directive reasons to do these things *because* of the blame. You do not get to dismiss your duty or reasons not to steal, nor do you get to dismiss duties or reasons to apologize, make amends, feel remorse, and so on. Consider one natural response to being hypocritically blamed for harming some third party: "I am going to apologize, but certainly not because *you*, of all people, have blamed

21 See Herstein, "Understanding Standing," 3116; and McDonough, "The Abuse of the Hypocrisy Charge in Politics."

22 Herstein, "Understanding Standing" and "Justifying Standing to Give Reasons." For accounts of standing that emphasize the importance of second-personal reasons and so are similar in important ways to Herstein's account, see Tognazzini, "On Losing One's Moral Voice"; and Piovarchy, "Hypocrisy, Standing to Blame and Second-Personal Authority."

23 It is less important for my purposes *why* hypocrisy or meddling gives the blamed party permission to dismiss blame. But briefly: Herstein's view is that there are norms against blaming hypocritically or meddlesomely, and allowing the blamed party to dismiss the blame is a kind of compensation once these norms have been violated ("Justifying Standing to Give Reasons," sec. 4.5).

me." The "because" here is clearly intended to be more than merely explanatory. The response means that I am not taking your blame to be a reason to apologize.

According to this view, failing to be moved to apologize, make amends, feel remorse, and so on for the directive reasons issued by blame amounts to dismissing that blame. I will argue that this is not the right way to understand what it is to dismiss blame. I rely in part on an intuitive understanding of when someone has or has not dismissed blame. I take it that we can pretty reliably detect *when* blame has been dismissed, and I make use of this in evaluating accounts of *what* such dismissal involves. One important guide for our judgments here is that dismissing *legitimate* blame—that is, blame that is fitting and for which the blamer has standing—is wrong, or at least wrongs the blamer, since we all have an interest in being able to hold one another to account.[24] So in a case in which the blamer does have standing, we can determine whether the blamed party has dismissed the blame by asking whether they have done something wrong, or at least wronged the blamer, in responding to the blame.[25]

Consider a case in which the blamer does have the standing to blame such that dismissing blame is not legitimate. If blame issues directive reasons that cannot be legitimately dismissed, then the blamed party should take them into account in their deliberations, and they should serve as the (or at least a) basis of the agent's actions or emotions. But it is not wrong and does not wrong the blamer for someone who has acted wrongly to feel remorse, apologize, make amends, and refrain from future wrongdoing not because they have been blamed but because of the ordinary moral reasons to do so. In fact, this will often be a *better* response, since the agent then is guided by the moral considerations rather than by the person blaming them. We expect good agents to apologize and feel guilty because they have done something wrong, not

24 I am not assuming that anytime we set back some of a person's interests, we necessarily wrong them. But I think it is clear that dismissing someone's blame when the blame is fitting and when they have standing is at least very often a way of wronging them. For discussion of the ways that people can be wronged by having their moral complaints dismissed or ignored, see Campbell, "Being Dismissed"; and Carbonell, "Social Constraints on Moral Address." For relevant discussion in the context of the standing to blame and of hypocrisy in particular, see Friedman, "How to Blame People Responsibly," 281; Wallace, "Hypocrisy, Moral Address, and the Equal Standing of Persons"; Fritz and Miller, "Hypocrisy and the Standing to Blame"; and Roadevin, "Hypocritical Blame, Fairness, and Standing."

25 Blame that is fitting and for which the blamer has standing might be objectionable in other ways, depending on how widely we understand "fitting." For example, the blame may be disproportionate, delivered in an overly aggressive tone, or at the wrong time. These kinds of considerations fall under what D. Justin Coates and Neal A. Tognazzini call *procedural* norms on blame ("The Nature and Ethics of Blame"). When these kinds of norms are violated, the blamed party will often be justified in *objecting* to the tone, timing, degree, etc. of the blame but not justified in *dismissing* the blame.

because they have been blamed. Intuitively, they can do so without thereby dismissing that blame.

One response is that where the blamer has standing, the agent's apologizing, feeling guilty, and so on should be *overdetermined*: the agent should do these things both for the ordinary reasons *and* for the directive reasons issued by the blame. This picture is plausible for other kinds of directives, such as requests. Suppose that you have independent reasons to take me to the airport: it would be a nice thing to do and would give us a chance to spend time together. My request for a ride to the airport adds to these reasons. You could sensibly cite my request as being among your reasons for taking me, and there will be cases in which the (directive reason issued by the) request is what "tips the scales" in favor of taking me. But notably, none of this seems to happen with blame. Suppose I wrong you and some third party blames me for it. It would be objectionable for me to cite as my reason for apologizing or feeling guilty that this third party blamed me, and it is hard to imagine a case where the blame is what tips the scales in favor of feeling guilty or apologizing. What seems appropriate in this case is apologizing and feeling guilty for the ordinary reasons to do so, and I can do this compatibly with accepting blame from the third party.[26]

Whatever accepting blame amounts to, contrary to what Herstein's view predicts, the blamed party can do it even while apologizing, feeling guilty, and so on for the ordinary moral reasons to do so, rather than for new directive reasons issued by the blame. Responding appropriately to your own wrongdoing is one thing, while responding appropriately to being blamed is another. This is not to say that accepting blame is compatible with not responding to it in any way, of course. The point is that one can accept blame without being moved to apologize, change one's ways, or feel guilty *because* one has been blamed.

To illustrate, imagine that the blamed party is already in the process of making amends, feeling remorse, changing their ways, and so on. Blame can still be appropriate, and the blamed party does not necessarily dismiss the blame, even if they do not suddenly add new directive reasons to their motivating reasons for doing these things. The blamed party may say something

26 In cases in which the person blaming you is the victim of your wrongdoing and they have standing, it is at least arguably better to be motivated to apologize both by the ordinary reasons to do so *and* because you have been blamed. This plausibly constitutes the appropriate recognition of their moral complaint. To preview, I hold that the relevant demand here—the one we dismiss when we dismiss blame—is for a second-personal expression of remorse to the blamer. A sincere apology to the person you have wronged constitutes a second-personal expression of remorse to them, and so where they are the person who has blamed you, you have both ordinary reasons to apologize and blame-specific reasons to apologize. My argument in the main text is just that apologizing to the victim because of a *third party's* blame is objectionable.

like: "I know; I just feel terrible. I'm on my way to apologize now." This does not seem to constitute dismissing blame, but there is no indication that the blame is playing a motivating role either in how the person feels or in their decision to apologize.

For another case, suppose that someone acts wrongly out of negligence caused by distraction rather than genuine ill will or lack of concern. Blaming them can make them aware of what they have done. Since the person is sensitive to the relevant moral considerations, they might feel guilty, get to work making amends, trying to make sure it never happens again, and so on, on the basis of those moral considerations to which they are sensitive, rather than for any new directive reasons arising from the blame itself. Even if the blame plays a causal role in their actions and emotions, it does not play a motivating role. This need not amount to wrongfully dismissing blame.

If we assume that accepting blame requires being moved by directive reasons to apologize, feel guilty, and change our ways, then it would be objectionable not to take them into account in a case in which the blamer has standing. The fact that it does *not* seem objectionable tells against thinking that accepting blame requires taking such reasons into account, whether we think blame in fact issues such reasons or not. Again, it does not follow that there is *nothing* that one should do in response to legitimate blame. Neither does it follow that any account of dismissing blame based on the dismissing of directive reasons, or demands more generally, should be rejected. The objection to this account turned on the fact that we already have many good reasons to do and feel the things that the posited directive reasons are reasons to do or feel and that someone can do these things for these reasons without thereby dismissing blame. The lesson is that we need to find something that is demanded of us once and because we have been blamed. This must be something we can do because of the blame, even if we go on to change our ways, apologize, and feel remorse for the independent moral reasons.

2.3. Second-Personal Responses

What do I do, then, if I accept the blame but then go on to do and feel the appropriate things for ordinary moral reasons? Many philosophers argue that blame demands some kind of *acknowledgment* or *recognition* from the blamed party.[27] This is a *second-personal* response to the person who is blaming you and, in particular, to their blaming you—they are the *recipient* of the response.

27 See Smith, "Control Responsibility, and Moral Assessment"; Martin, "Owning Up and Lowering Down"; Friedman, "How to Blame People Responsibly"; Macnamara, "'Screw You!' and 'Thank You!'"; Fricker, "What's the Point of Blame?"; and Edwards, "Standing to Hold Responsible."

It is thus something that we can only do, and so only be under a demand to do, once we have been blamed. Unlike being moved to apologize, feel remorse, or change one's ways by ordinary moral reasons rather than by the blame, failing to acknowledge the blame does plausibly amount to dismissing that blame. So it takes on board the lesson from the discussion of Herstein's account. The important question is what this acknowledgment amounts to.

Edwards offers a view of dismissing blame that emphasizes the importance of acknowledging and engaging with the blame. He holds that blaming someone involves an accusation of wrongdoing, which itself involves a demand for some fitting *content-sensitive* response to that accusation, that is, one that "engages with the accusations *on their merits*" by accepting or denying the content.[28] One way of doing this is to deny the content, that is, deny wrongdoing. There are also various ways of accepting the accusation, where different kinds of accusations (e.g., condemnation versus mild criticism) demand different kinds of accepting responses. These include "expressions of remorse, or acts of repentance." Dismissing blame, in Edwards's view, "is to refuse to accede to this demand" and to instead offer a content-*insensitive* response, such as a charge of hypocrisy, or perhaps no response at all. This is to "implicitly deny that a content-sensitive response is owed" to the blamer (449).

Edwards contrasts acceptance of the *content of an accusation* with dismissing the blame that makes that accusation. But we can and frequently do accept the content of the accusation involved in blame, and we may even tell the blamer that we do, even when we dismiss blame: "Sure, I shouldn't have done it, but who are *you* to blame me for it?" Edwards can accept this, since denying that a content-sensitive response is owed to the blamer is compatible with accepting the content of the accusation. But we should emphasize that the question crucial for determining whether we have dismissed blame is not whether we engage with the *accusation* by accepting or denying its content (447) but whether we engage with the *accuser* in the right way.

I agree with Edwards's focus on expressing remorse. But we should more explicitly highlight that the demand is for a *second-personal* response to the blamer. An expression of remorse to the person we have wronged through a sincere apology, when they are not the person blaming us, is fully compatible with dismissing blame. So just expressing remorse cannot be sufficient for accepting blame. Rather, the expression of remorse must be *to* the blamer; they must be the recipient of the expression of remorse.

28 Edwards, "Standing to Hold Responsible," 447. Citations to Edwards's paper in the next few paragraphs will be parenthetical.

Macnamara offers an account of acknowledging blame that highlights its status as a second-personal response to being blamed. She argues that expressed blame seeks (i) recognition that the blamer has correctly recognized you as having done something wrong, and thus (ii) recognizing yourself as a wrongdoer, and (iii) the *expression* of this recognition. To recognize that the blamer has correctly recognized you, and thus to recognize yourself, as a wrongdoer involves feeling emotions such as guilt and remorse. To express this recognition is to give voice to these emotions by, according to Macnamara, "apologizing and making amends if necessary."[29]

Insofar as recognition is a distinctively second-personal response, this account is on the right track. But even when we legitimately dismiss blame, we can and should do all of the things Macnamara mentions. Suppose that you have acted wrongly and someone blames you, but their blame is hypocritical, and so you legitimately dismiss it. Still, you can and should (i) recognize that the blamer has correctly identified you as a wrongdoer, since you do not *deny* wrongdoing, (ii) experience remorse, since this is called for when you know that you acted wrongly, and (iii) apologize or make amends, especially (though not only, I believe) if the person hypocritically blaming you is not the person you wronged. You can respond to the hypocritical blame by saying something like: "Look, I know I shouldn't have done it, and I feel bad about this. In fact, I'm on my way to apologize right now. But who are *you*, of all people, to blame me for this?"

The view I prefer, and will develop in the next section, follows Macnamara in holding that the demand we dismiss is one for a *second-personal* response to the blamer as a response to their blame. But it moves beyond responses we should have anyway, such as apologizing, feeling guilty, or making amends. In my view, Edwards is correct to identify *expressing remorse* as the response that blame demands and Macnamara is correct to emphasize that the response demanded is a distinctively second-personal one. I develop this thought and argue that this gives us an attractive account of what we dismiss when we dismiss blame.

3. EXPRESSING REMORSE

Macnamara notes the importance of *giving voice* to our guilt or remorse when we are blamed. It is not enough to accept blame just to admit that you have these emotions or to apologize to the person you have wronged and make

29 Macnamara, "'Screw You!' and 'Thank You!,'" 909. Macnamara is here drawing on Rebecca Kukla and Mark Lance on *recognitives* (Kukla and Lance, *"Yo!" and "Lo!"*). Adrienne Martin similarly claims that resentment or blame asks the wrongdoer to (i) take ownership of the wrongful deed, (ii) regret it, and (iii) make moral repair ("Owning Up and Lowering Down," 545).

amends—you can do these things consistently with dismissing blame, especially if the person blaming you is not the person to whom you owe an apology. But there are other ways of giving voice to these emotions. Since what we are after is a direct, second-personal response to being blamed, it is natural to turn to apology, as Macnamara does. Apology is the paradigmatic second-personal response to our own wrongdoing. But since the person blaming you may not be the person to whom you owe an apology, this is not quite right, or at least it will not cover all cases.

Still, we can make progress by thinking about what is involved in apologizing and, especially, about the difference between sincere and insincere apologies. Sincere apologies express remorse that the apologizer feels, while insincere apologies often ring hollow because the apologizer does not actually feel remorse. But as Tierney observes, even if you truly claim that you are remorseful in the process of apologizing, the apology may nevertheless seem insincere or cold.[30] The recipient may tell you to "say it like you mean it." A sincere apology should allow the recipient to see or witness your remorse rather than merely tell them that you feel it. Martin stresses the "performative element of apology," which displays "regretful ownership" of the wrongdoing.[31] Owning up is not enough, even if you truly *say* that you are remorseful or regretful. The apologizer needs to performatively express or display their remorse to the recipient.[32] Several philosophers defend views of apology that develop this thought, emphasizing the importance of communicating to the recipient of an apology that you are giving them the power to decide whether to forgive you and move on from the wrongdoing. This is to put the recipient in charge of, as Bovens puts it, "restoring [your] moral stature."[33] Communicating this typically involves a humble apology that clearly expresses remorse to the recipient.

Even if the blamer does not demand an apology, since your wrongdoing may not have affected her, in blaming you, she can still demand a performative expression of the kind of remorse characteristic of sincere apology. Perhaps she also

30 Tierney, "Don't Suffer in Silence," sec. 4.b.

31 Martin, "Owning Up and Lowering Down," 547.

32 Hannah Tierney argues on various ethical grounds for the importance of expressing your self-blame, via an expression of guilt, to those you have wronged ("Don't Suffer in Silence"). I am suggesting that at least some instances of blame may involve a demand or expectation for this kind of expression of guilt or remorse to the blamer, even if the blamer is not the one you have wronged. Compare Tierney's discussion of publicly expressing guilt when the wrongdoer is unreachable in section 4.c.

33 Bovens, "Apologies," 231. On the inadequacy of an "interview apology," where you inform the victim that you feel bad, will not do it again, etc. by coolly answering a series of questions such as "Do you regret it?," "Will you do it again?," etc., see Helmreich, "The Apologetic Stance," 79–80. See also Martin, "Owning Up and Lowering Down."

demands that you sincerely apologize and so express remorse to the victim(s) of your wrongdoing. In general, blame might involve several demands. Relatedly, as Macnamara emphasizes, we should distinguish between what would satisfy the demands involved in blaming and what would satisfy an individual blamer.[34] Many blamers will rightly care more about the wrongdoer apologizing to the victim or improving their future behavior than about whether the blamer expresses remorse to them. But as we have seen, complying with demands to do these things is not necessary for accepting blame. The central claim here is that in blaming you, the blamer (also) demands that you express remorse *to her*, that is, to the blamer, even if she does not demand that you apologize to her.[35]

Complying with this demand seems to me a good candidate for what is necessary for accepting hostile blame involving indignation or resentment. Expressing remorse to the blamer is a way of acknowledging that they are right that you have acted wrongly and showing that you are pained by your behavior. Many authors have taken angry blaming attitudes such as resentment and indignation to involve a (usually indeterminate) desire for some kind of suffering (broadly construed) on the part of the wrongdoer caused by recognition of their wrongdoing. An expression of remorse to the blamer satisfies this desire.[36] Consider Rosen's remark that "the wrongdoer who responds to outward blame with a sincere and cheerful promise to do better next time but without a hint of guilt or remorse palpably frustrates the desire implicit in resentment" and, we could add, indignation.[37]

If Rosen is correct, then accepting hostile blame plausibly requires an expression of remorse. For all he has said here, this remorse might be expressed only to the victim and not to the blamer. But as I have argued, expressing remorse only to the *victim* is consistent with dismissing blame. So to count as accepting blame, this remorse needs to be expressed to the blamer. You can be remorseful, and even tell the blamer that you are remorseful, without *letting her in on* that remorse: "Yes, of course I feel bad, but I'm not going to sit here and take this from you of all people." In calling this a second-personal expression of remorse to the blamer, I mean to distinguish a case in which the blamer is

34 Macnamara, "'Screw You!' and 'Thank You!,'" 896, 899.

35 I will continue to use "remorse" here, since it is commonly used in the literature, rather than focus on distinctions between remorse and other emotions of negative self-assessment (cf. Taylor, *Pride, Shame, and Guilt*). But perhaps "contrition" would be better; Bero emphasizes the importance of expressions of contrition for moral repair ("Holding Responsible and Taking Responsibility").

36 Nussbaum objects to anger on the basis of this kind of desire for suffering (*Anger and Forgiveness*).

37 Rosen, "The Alethic Conception of Moral Responsibility," 82–83.

the *recipient* of the expression of remorse from the case in which the blamer is merely (part of) the *audience* of an expression of remorse. If the blamer happens to see you sincerely and remorsefully apologize to the person you have wronged, this need not constitute accepting blame from the blamer. Accepting hostile blame involves a second-personal expression of remorse *to* the blamer as recipient in the way characteristic of a sincere apology, even if you are not apologizing to her.

My proposal is that this is what we can dismiss when we can dismiss blame: a demand that we express our remorse to the blamer. This account of dismissing blame has several good features. First, it is appropriately localized to the *blamer*: just because you do not have to express your remorse to the hypocrite blaming you does not mean you do not have to express it to someone else who blames you non-hypocritically. The account is also localized to the *blame*: the demand to express remorse to the blamer is issued by the blame itself. Doing so is a response to being blamed and so a way of acknowledging the blame. Third, refraining from expressing remorse to the blamer is consistent with doing all the things that you should be doing independently of being blamed: admitting that you have acted wrongly, feeling remorse, and expressing that remorse *to the person you have wronged* through a sufficiently sincere apology.[38] As we have seen, dismissing blame is consistent with doing these things, as well. What you dismiss is the demand for the second-personal response of expressing your remorse *to* the blamer.

Can we say more about what is involved in expressing remorse to the blamer? I suspect it is highly dependent on features of the context, including the personalities and social identities of the blamer and blamed party, the relationship between them, general cultural norms, and the nature of the wrongdoing in question. This makes it difficult to say anything very general beyond the suggestive remarks I have made so far. But some familiar ways of doing so

38 What if you wrong someone who regularly wrongs you or others in the same way, but they blame you anyway? Here my intuitions are not clear; others have expressed similar ambivalence (see also Smilansky, "The Paradox of Moral Complaint"). Since their blame is hypocritical, it seems that you can dismiss it. But you have wronged them, and so seem to owe them an apology simply on that basis. If you still owe them an apology, as seems plausible in at least some cases, in what sense can you dismiss their blame? I am inclined to think that you still owe them a sincere apology, and since this involves an expression of remorse, you do owe them this, as well. But perhaps it is inappropriate for them to *demand* this expression of remorse, given the hypocrisy involved in doing so. So in a case in which the person to whom you owe an apology blames you and is not hypocritical in doing so, your owing them a sincere expression of remorse is overdetermined: you owe them a sincere apology, which expresses remorse, in virtue of wronging them, and, in addition, an expression of remorse in response to their blame. Of course, since a sincere apology to someone is a paradigmatic way to express remorse to them, the expressions of remorse need not be distinct.

include downcast eyes, averted gaze, and, in the extreme case, groveling or (usually figuratively) throwing oneself at the feet of the blamer.[39] Of course, sincere claims that you are remorseful, apologies, paying compensation, and so on also show remorse. But it is the expression of remorse *to* the blamer that can be dismissed when you can dismiss blame.

In arguing for this account of dismissing blame, I argued that accepting hostile blame requires complying with the demand for a second-personal expression of remorse to the blamer. It is compatible with my claim about dismissing blame, though, that some kinds or instances of blame do not, in fact, involve this demand. I think this is especially plausible for gentler, nonhostile kinds of blame or criticism. But importantly, these kinds of criticism are not typically apt for dismissal on the basis of lack of standing. If I point out to you that you have mistreated some third party in a constructive or understanding way, perhaps even admitting my own recent mistreatment of someone, it would be inappropriate for you to respond by dismissing this criticism as hypocritical, for example.[40]

This brings out one way in which dismissing blame or criticism can go wrong besides the more obvious case in which the blamer really does have standing. Targets of criticism often inappropriately dismiss that criticism on the basis of lack of standing when, in fact, the critic was not engaged in hostile blame. Consider constructive criticism that certain choices we make—eating meat or taking short-haul flights, for example—are morally questionable, given the climate crisis. It is easy to imagine such criticism being met with angry dismissal: "Who are you to scold me about this? Didn't you fly to Italy just last summer?" And it is equally easy to imagine a case in which the critic could appropriately respond by saying something like, "Whoa, look, I'm not trying to get in your face about this—we all have a lot of work to do." The dismissive response is not apt in this case, and, in my view, that is because the criticism is nonhostile and does not involve the demand for an expression of remorse to the critic.

A final complication involves blame that is not expressed to the blamed party. As I noted at the beginning of the paper, blame that is expressed to a third party can be dismissed by that third party and blame that is kept private can, if it is discovered, be dismissed. But if my account is right, then this blame must therefore involve a demand that the blamed party express remorse to the blamer. It is hard to see how it could do this if the blame is not even expressed

39 All of these can be faked, of course, but that does not cast doubt on the claim that they are familiar ways of genuinely expressing remorse.

40 For discussion of this point, see, e.g., Dworkin, "Morally Speaking," 184; Cohen, "Ways of Silencing Critics," 139; Rivera-López, "The Fragility of Our Moral Standing to Blame," 345; Isserow and Klein, "Hypocrisy and Moral Authority," 199; and King, "Skepticism about the Standing to Blame," 1437–38.

to the blamed party. Though I think it is sensible to treat direct, expressed blame—and its dismissal—as the paradigm case, I do need to say something about how to extend what I have said here to these kinds of cases.

This problem will face anyone who takes blame to involve demands, as I have done, and fully addressing this issue on behalf of this understanding of blame would take more space than I have here. But briefly, I think it is plausible that blame can involve demands on the blamed party even if it is not expressed to them. We can privately make demands of people or make demands of them to others. Think of saying to yourself or to another colleague, "Bob had better not reheat fish in the department microwave again today." This is plausibly making a demand of Bob, even though you are not expressing it to Bob. Likewise, then, blaming Bob for doing this can be understood as involving a demand that he expresses remorse to you for his uncivil behavior, even if you do not express that demand to Bob. If you lack standing to blame Bob for this—for example, if you regularly reheat fish—then your colleagues or Bob himself, if he finds out about your blame, can dismiss the demand. Of course, if Bob does not know about your demand, he is unlikely to comply with it. But that is true of demands we keep to ourselves generally. This is part of why keeping demands and, especially, blame to ourselves can lead to increased resentment and frustration.

4. Dismissing Blame and the Standing to Blame

Before closing, I will briefly connect this account of what it is to dismiss blame to the standing to blame. Since part of what it is to lack standing to blame is for the target of the blame to be justified in dismissing that person's blame, we can approach issues about standing from a new direction. In addition to asking whether some condition undermines standing directly, we can ask whether the condition justifies dismissing blame. If so, that is evidence that the condition does undermine standing.[41] On my account, we can ask whether and, if so, why this condition justifies dismissing a demand for a second-personal expression of remorse to the blamer. I will suggest that this approach to theorizing about standing allows us to see why two apparently unrelated conditions have the same effect of undermining standing to blame. Fully investigating this would take more space than I have here, but this should show the fruitfulness of thinking about dismissing blame.

Two important ways someone can lack standing to blame are: (i) the would-be blamer is themselves guilty of violating the norm, and (ii) the norm violation is none of the blamer's business. The former is most widely

41 On the *deflection test* for standing, see Edlich, "What about the Victim?," 213.

discussed in the recent literature, often under the heading of *hypocritical* blame. Accounts of why past wrongdoing undermines standing to blame typically do not, and are not intended to, generalize to the second kind of standingless blame, which is sometimes called *meddlesome* blame. Though your being guilty of the same kind of wrongdoing and the wrongdoing in question being none of your business are very different considerations, it is striking that both have the same effect of undermining standing and that both license the same kind of dismissive response.

On my account, dismissing blame is dismissing a demand for a second-personal expression of remorse to the blamer of the sort that is characteristic of a sincere apology. As we have seen, a popular thought is that sincere, remorseful apologies communicate that the apologizer is putting the power of, as Bovens says, "restoring [their] moral stature" in the hands of the recipient.[42] Assuming that (most) wrongdoing does not result in any reduction in fundamental worth, the most plausible way to understand this is in terms of *forgiveness*. Apologies put the power in the recipient's hands to forgive and so, as Walker says, make the "*morally reparative* decision to release himself or herself from the position of grievance and reproach, and to release the wrongdoer from open-ended (but not necessarily all other) demands for satisfaction."[43] According to this view, though the forgiver may still demand various kinds of compensation and may be unwilling to act as if the wrongdoing never happened, she commits to refrain from treating the original wrongdoing as a reason for further resentment, mistrust, and punishment.

Focusing on apology and forgiveness is too narrow for thinking about blame and proper responses to blame, since we can blame those who wrong third parties. In these cases, the blamed party does not owe *us* an apology and it is not *our* place to forgive them.[44] Still, according to my view, a second-personal expression of remorse in response to being blamed has important features in common with a sincere, remorseful apology. Letting the blamer in on our remorse in this way plausibly makes us subordinate to them in a similar way, putting the power to decide whether to refrain from treating our wrongdoing as a reason for further indignation, reproach, and punishment. I will now suggest

42 Bovens, "Apologies," 231.

43 Walker, *Moral Repair*, 153. See also page 157, where Walker discusses the views of Uma Narayan ("Forgiveness, Moral Reassessment, and Reconciliation") and Avishai Margalit (*The Ethics of Memory*).

44 For recent discussion of a related asymmetry between having standing to blame and having standing to forgive, see Fritz and Miller, "A Standing Asymmetry between Blame and Forgiveness."

that it makes sense to dismiss a demand to give the blamer this power in both kinds of cases in which they intuitively lack standing.

First, if the blamer is themselves guilty of the same kind of wrongdoing, then we can reasonably dismiss a demand to take up a subordinate position relative to them and put the power to restore our moral stature in their hands. This is because, due to their own similar wrongdoing, they have not earned this kind of elevated position. They themselves need to be welcomed back into the moral fold and so are not in a place to welcome us back. This approach fits particularly well with one candidate view of standing, according to which having standing to blame requires being *better*, in the relevant respect, than the person you are blaming.[45] If dismissing blame involves dismissing a demand to take up a subordinate position, treating the blamer as better in an important sense, then we can see why we would be justified in doing so when they are not better. This account of standing is controversial, of course, but it is notable that this independently motivated account of what it is to dismiss blame provides at least *prima facie* support for it.[46]

This explanation can also plausibly be made to fit with the other main candidate account—and probably the most popular one—of how past wrongdoing undermines standing, according to which such blamers lack standing because they treat themselves as above the law, morally speaking.[47] At least many such blamers will make demands that the blamed party give the blamer the power to restore their moral stature while being unwilling to accede to such demands from others, stubbornly maintaining that they have nothing to be remorseful for and do not need their moral stature to be restored. Plausibly, this kind of self-preferential attitude undermines their standing to issue the relevant demands.

Second, consider meddlesome blame, which we often dismiss, saying something along the lines of "Who are *you* to blame me for this? Mind your own business!" This kind of dismissal of blame makes sense in cases in which the wrongdoing in question is in the context of some private relationship, for example, within a family. Obviously, there are wrongs that take place within

45 See, e.g., Dworkin, "Morally Speaking"; Rivera-López, "The Fragility of Our Moral Standing to Blame"; and Todd, "Let's See You Do Better." Jessica Isserow and Colin Klein do not explicitly endorse this view, but some of what they say can be taken as support for it ("Hypocrisy and Moral Authority").

46 I defend this view of standing, drawing on the account of dismissing blame developed here, in my "Explaining Loss of Standing to Blame."

47 See, e.g., Wallace, "Hypocrisy, Moral Address, and the Equal Standing of Persons"; Fritz and Miller, "Hypocrisy and the Standing to Blame" and "The Unique Badness of Hypocritical Blame"; Roadevin, "Hypocritical Blame, Fairness, and Standing"; and Piovarchy, "Hypocrisy, Standing to Blame and Second-Personal Authority."

the context of a family that genuinely are everyone's business, but plausibly there are spheres of privacy such that wrongdoing within that sphere is not the business of other people.[48] If blame involves a demand for a second-personal expression of remorse, and complying with this demand involves taking up a subordinate position relative to the blamer such that the blamer is given the power to decide whether to welcome the wrongdoer back into the moral fold, we can see why dismissing meddlesome blame is warranted. In short, it is because the blamer is not part of the relevant fold. It is not that they have alienated themselves from it by acting immorally but rather that they were never part of it to begin with. Just as in the case of hypocritical blame, the target of such blame can be justified in dismissing a demand for a second-personal expression of remorse to the blamer. The private wrongdoing at issue in these cases is not properly treated as a reason for indignation, mistrust, or punishment for people outside of the relevant sphere, and so it is inappropriate for these people to demand a display of remorse that gives them power to decide to refrain from treating it as a reason for these reactions.

There is, of course, much more to say about both hypocritical blame and meddlesome blame, and I have only been able to sketch the explanations that my account of dismissing blame suggests for why such blame can be dismissed. Still, I think this illustrates the fruitfulness of approaching the standing to blame from the perspective of the blamed party and asking what kinds of responses are appropriate and why.[49]

University of St Andrews
js280@st-andrews.ac.uk

REFERENCES

Bell, Macalester. "The Standing to Blame: A Critique." In *Blame: Its Nature and Norms*, edited by D. Justin Coates and Neal A. Tognazzini, 263–81. New

48 See, e.g., Smith, "On Being Responsible and Holding Responsible"; Radzik, "On the Virtue of Minding Our Own Business"; and Seim, "The Standing to Blame and Meddling."

49 For very helpful discussion of material included in this paper, thanks to Zoë Johnson King, Julia Driver, Joe Bowen, Jessica Brown, Emmy Feamster, Adam Etinson, Julia Staffel, Erik Encarnacion, Ben Lennertz, Nathan Howard, Coleen Macnamara, Patrick Todd, Maggie O'Brien, Hannah Tierney, Dana Nelkin, Shyam Nair, Per-Erik Milam, David Shoemaker, Andrew Ma, anonymous referees, and audiences at Stirling, St Andrews, Glasgow, the Cambridge Moral Sciences Club, and the Univeristy of Texas at Austin/St Andrews Workshop on Blame and Responsibility.

York: Oxford University Press, 2013.
Bero, Stephen. "Holding Responsible and Taking Responsibility." *Law and Philosophy* 39, no. 3 (June 2020): 263–96.
Bovens, Luc. "Apologies." *Proceedings of the Aristotelian Society* 108, nos. 1–3 (October 2008): 219–39.
Campbell, Sue. "Being Dismissed: The Politics of Emotional Expression." *Hypatia* 9, no. 4 (Summer 1994): 46–65.
Carbonell, Vanessa. "Social Constraints on Moral Address." *Philosophy and Phenomenological Research* 98, no. 1 (January 2019): 167–89.
Coates, D. Justin, and Neal A. Tognazzini. "The Nature and Ethics of Blame." *Philosophy Compass* 7, no. 3 (March 2012): 197–207.
Cohen, G. A. "Casting the First Stone: Who Can, and Who Can't, Condemn the Terrorists." In *Finding Oneself in the Other*, 115–33. Oxford: Oxford University Press, 2013.
———. "Ways of Silencing Critics." In *Finding Oneself in the Other*, 134–42. Oxford: Oxford University Press, 2013.
Collins, Stephanie. "Filling Collective Duty Gaps." *Journal of Philosophy* 114, no. 11 (November 2017): 573–91.
Darwall, Stephen. *The Second-Person Standpoint*. Oxford: Oxford University Press, 2006.
Dover, Daniela. "Criticism as Conversation." *Philosophical Perspectives* 33, no. 1 (December 2019): 26–61.
———. "The Walk and the Talk." *Philosophical Review* 128, no. 4 (October 2019): 387–422.
Driver, Julia. "The Suberogatory." *Australasian Journal of Philosophy* 70, no. 3 (September 1992): 286–95.
Duff, R. A. "Blame, Moral Standing, and the Legitimacy of the Criminal Trial." *Ratio* 23, no. 2 (June 2010): 123–40.
Dworkin, Gerald. "Morally Speaking." In *Reasoning Practically*, edited by Edna Ullmann-Margalit, 182–88. Oxford: Oxford University Press, 2000.
Edlich, Alexander. "What about the Victim? Neglected Dimensions of the Standing to Blame." *Journal of Ethics* 26, no. 2 (June 2022): 209–28.
Edwards, James. "Standing to Hold Responsible." *Journal of Moral Philosophy* 16, no. 4 (August 2019): 437–62.
Fricker, Miranda. "What's the Point of Blame? A Paradigm Based Explanation." *Noûs* 50, no. 1 (March 2016): 165–83.
Friedman, Marilyn. "How to Blame People Responsibly." *Journal of Value Inquiry* 47, no. 3 (September 2013): 271–84.
Fritz, Kyle G., and Daniel J. Miller. "Hypocrisy and the Standing to Blame." *Pacific Philosophical Quarterly* 99, no. 1 (March 2018): 118–39.

———. "A Standing Asymmetry between Blame and Forgiveness." *Ethics* 132, no. 4 (July 2022): 759–86.

———. "The Unique Badness of Hypocritical Blame." *Ergo* 6, no. 19 (2019–2020): 545–69.

Helmreich, Jeffrey S. "The Apologetic Stance." *Philosophy and Public Affairs* 43, no. 2 (Spring 2015): 75–108.

Herstein, Ori. "Justifying Standing to Give Reasons: Hypocrisy, Minding Your Own Business, and Knowing One's Place." *Philosophers' Imprint* 20, art. 7 (2020): 1–18.

———. "Understanding Standing: Permission to Deflect Reasons." *Philosophical Studies* 174, no. 12 (December 2017): 3109–32.

Hieronymi, Pamela. "The Force and Fairness of Blame." *Philosophical Perspectives* 18, no. 1 (2004): 115–48.

Isserow, Jessica, and Colin Klein. "Hypocrisy and Moral Authority." *Journal of Ethics and Social Philosophy* 12, no. 2 (2017): 191–222.

King, Matt. "Skepticism about the Standing to Blame." In *Oxford Studies in Agency and Responsibility*, vol. 6, edited by David Shoemaker, 265–88. Oxford: Oxford University Press, 2019.

Kukla, Rebecca, and Mark Lance. *"Yo!" and "Lo!": The Pragmatic Topography of the Space of Reasons*. Cambridge, MA: Harvard University Press, 2009.

Lippert-Rasmussen, Kasper. *Relational Egalitarianism: Living as Equals*. Cambridge: Cambridge University Press, 2018.

Macnamara, Coleen. "'Screw You!' and 'Thank You!'" *Philosophical Studies* 165, no. 3 (September 2013): 893–914.

———. "Taking Demands Out of Blame." In *Blame: Its Nature and Norms*, edited by D. Justin Coates and Neal A. Tognazzini, 141–61. New York: Oxford University Press, 2013.

Margalit, Avishai. *The Ethics of Memory*. Cambridge, MA: Harvard University Press, 2002.

Martin, Adrienne. "Owning Up and Lowering Down: The Power of Apology." *The Journal of Philosophy* 107, no. 10 (October 2010): 534–53.

McDonough, Richard. "The Abuse of the Hypocrisy Charge in Politics." *Public Affairs Quarterly* 23, no. 4 (October 2009): 287–307.

McKenna, Michael. *Conversation and Responsibility*. Oxford: Oxford University Press, 2012.

Narayan, Uma. "Forgiveness, Moral Reassessment, and Reconciliation." In *Explorations of Value*, edited by T. Magnell, 169–78. Amsterdam: Brill, 1997.

Nussbaum, Martha. *Anger and Forgiveness*. Oxford: Oxford University Press, 2016.

O'Brien, Maggie, and Alexandra Whelan. "Hypocrisy in Politics." *Ergo* 9, no. 63

(2022): 1692–714.
Piovarchy, Adam. "Hypocrisy, Standing to Blame and Second-Personal Authority." *Pacific Philosophical Quarterly* 101, no. 4 (December 2020): 603–27.
Radzik, Linda. "On the Virtue of Minding Our Own Business." *Journal of Value Inquiry* 46, no. 2 (June 2012): 173–82.
Rivera-López, Eduardo. "The Fragility of Our Moral Standing to Blame." *Ethical Perspectives* 24, no. 3 (2017): 333–61.
Roadevin, Cristina. "Hypocritical Blame, Fairness, and Standing." *Metaphilosophy* 49, nos. 1–2 (January 2018): 137–52.
Rosen, Gideon. "The Alethic Conception of Moral Responsibility." In *The Nature of Moral Responsibility: New Essays*, edited by Randolph Clarke, Michael McKenna, and Angela M. Smith, 65–88. Oxford: Oxford University Press, 2015.
Rossi, Benjamin. "The Commitment Account of Hypocrisy." *Ethical Theory and Moral Practice* 21, no. 3 (June 2018): 553–67.
Scanlon, T. M. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, MA: Harvard University Press, 2008.
Seim, Maria. "The Standing to Blame and Meddling." *Teorema* 38, no. 2 (2019): 7–26.
Sher, George. *In Praise of Blame*. Oxford: Oxford University Press, 2006.
Shoemaker, David. "Moral Address, Moral Responsibility, and the Boundaries of the Moral Community." *Ethics* 118, no. 1 (October 2007): 70–108.
Smilansky, Saul. "The Paradox of Moral Complaint." *Utilitas* 18, no. 3 (September 2006): 284–90.
Smith, Angela. "Control, Responsibility, and Moral Assessment." *Philosophical Studies* 138, no. 3 (April 2008): 367–92.
———. "On Being Responsible and Holding Responsible." *Journal of Ethics* 11, no. 4 (December 2007): 465–84.
Snedegar, Justin. "Explaining Loss of Standing to Blame." *Journal of Moral Philosophy* (forthcoming).
Strawson, P. F. "Freedom and Resentment." *Proceedings of the British Academy* 48 (1962): 1–25.
Taylor, Gabriele. *Pride, Shame, and Guilt: Emotions of Self-Assessment*. Oxford: Oxford University Press, 1985.
Tierney, Hannah. "Don't Suffer in Silence: A Self-Help Guide to Self-Blame." In *Self-Blame and Moral Responsibility*, edited by Andreas Brekke Carlsson, 117–33. Cambridge: Cambridge University Press, 2022.
Todd, Patrick. "Let's See You Do Better." *Ergo* (forthcoming).
———. "A Unified Account of the Moral Standing to Blame." *Noûs* 53, no. 2 (June 2019): 347–74.

Tognazzini, Neal. "On Losing One's Moral Voice" (unpublished manuscript). https://philpapers.org/archive/TOGOLO.pdf.

Walker, Margaret Urban. *Moral Repair: Reconstructing Moral Relations after Wrongdoing*. Cambridge: Cambridge University Press, 2006.

Wallace, R. J. "Emotions, Expectations, and Responsibility." In *Free Will and Reactive Attitudes*, edited by Michael McKenna and Paul Russell, 157–85. Farnham: Ashgate, 2008.

———. "Hypocrisy, Moral Address, and the Equal Standing of Persons." *Philosophy and Public Affairs* 38, no. 4 (2010): 307–41.

———. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press, 1996.

Watson, Gary. "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme." In *Responsibility, Character, and the Emotions*, edited by Ferdinand Schoeman, 256–86. New York: Cambridge University Press, 1988.

Wolf, Susan. *Freedom within Reason*. Oxford: Oxford University Press, 1990.

Journal of Ethics and Social Philosophy
VOL. 26, NO. 3 · FEBRUARY 2024
https://doi.org/10.26556/jesp.v26i3.2370

THE PROBLEM OF BASIC EQUALITY

A CONSTRUCTIVE CRITIQUE

Nikolas Kirby

IT IS COMMON to assume that all (or nearly all) human beings are one another's *equals* in some basic moral sense. Indeed, some have argued that this assumption forms an "egalitarian plateau" for contemporary political philosophy: a foundational, shared premise from which all plausible theories must proceed.[1] However, there is a problem with this assumption of *basic equality*, indeed arguably "one of the most profound problems of moral philosophy."[2] There must be *some* explanation as to why all (or nearly all) humans are equals with one another, but not with other beings. Most plausibly, this must, in part, be because we possess some nonnormative property or properties that ground our equality. Yet when we aim to isolate such nonnormative properties, we find that they tend to come in differing degrees across the human population. But this raises the question: If having some of a nonnormative property (like "rationality") is supposed to be so important that it grounds having some of a basic moral property (like "worth"), then why does having more of that nonnormative property not also ground having more of that basic moral property? There is a seemingly unavoidable "pressure of reason" to conclude the exact *opposite* of basic equality: we must be one another's *un*equals.[3]

This paper offers a targeted five-point critique of the current debate with respect to this problem, with the aim of laying the foundations for a new strategy. First, it claims that the debate should be refocused away from any particular concept(ion) of basic equality to a more agnostic proposition about the possibility of establishing equality in any basic moral property (section 1). Second, it rearticulates the problem in terms of grounding relations rather than supervenience (section 2). Third, it focuses on the current dominant approach to solving this problem, the nonscalar property strategy. It argues that proponents

1 See Kymlicka, *Contemporary Political Philosophy*, 4; Dworkin, *Sovereign Virtue*, 11.

2 Christiano, *The Constitution of Equality*, 17.

3 Arneson, "Basic Equality," 36.

of this approach have failed to properly distinguish between two different nonscalar properties defined in terms of scalar properties: *range properties* and *bare properties* (section 3). Once such properties have been disambiguated, it becomes apparent that current proponents of the nonscalar strategy have failed to solve the pressure of reason problem: they have merely shifted it (section 4). However, I end by arguing that this critique directs our attention to a possible alternative strategy—that is, grounding our equality on a *relative property* (section 5). I illustrate this strategy via a resuscitation of a neglected line of thought from the first contemporary analysis of the problem of basic equality by Herbert Spiegelberg.[4] While not seeking to offer a full-throated defense of a solution to the problem of basic equality, this illustration aims to build the case for moving beyond a nonscalar range or bare property strategy to explore the possibilities of equality in relative properties instead.

1. DEFINING BASIC EQUALITY

Philosophers who are nominally addressing the same *problem* of basic equality tend to formulate the *concept* of basic equality in different ways. They speak of "equal worth," "equal consideration," "equal intrinsic value," "equal respect," "fundamental equal moral status," "equal standing," "entitlement to equal amounts of something," "the principles of justice which require equal basic rights be assigned to all persons … equal justice," "equal status … [implying] that moral principles in some way assert that [each] has a right to something, may be owed something, may deserve something, or that it ought to receive a certain good," and many other formulations.[5]

Are all of these formulations of basic equality equivalent? Are all these philosophers aiming to solve the same problem? Some argue (or more often appear to simply assume) that despite the apparent differences in the formulations of basic equality, they are synonymous, related by implication, or derivable (*qua* "conceptions") from a common claim that we are equals in one of the more abstract properties (*qua* "concept").[6] Others, however, stress the differences

4 Spiegelberg, "A Defense of Human Equality."

5 Vlastos, "Justice and Equality," 43; Benn, "Egalitarianism and the Equal Consideration of Interests"; Bedau, "Egalitarianism and the Idea of Equality," 17; Williams, *Problems*, 236; Arneson, "Basic Equality," 30; Sher, *Me, You, Us*, 32; Carter, "Respect and the Basis of Equality," 539; Rawls, *A Theory of Justice*, 442; Christiano, "Rationality, Equal Status, and Egalitarianism," 54.

6 Pojman and Westmoreland, *Equality*, 1; Dworkin, *Taking Rights Seriously*, 272–73; Sher, *Me, You, Us*, 31; Kymlicka, *Contemporary Political Philosophy*, 4; Miller, *National Responsibility and Global Justice*, 28.

and incompatibilities between them. They argue that beneath the thin veneer of a common commitment to sharing equality in *some* sense across political philosophy lies fundamental disagreement about *which particular* sense.[7]

Do we, then, need to resolve this debate or pick our own preferred formulation of basic equality in order to then formulate its *problem*? No. Indeed, there are distinct advantages to focusing on the following functionally defined proposition instead:

> *Basic Equality*: All (or nearly all) humans are one another's equals in their possession of some basic moral property.

In this context, "basic moral property" is simply a placeholder for *any* property that will do the kind of theoretical work we expect of a concept(ion) of basic equality. Thus, a property is *moral* if, depending on one's most primitive moral concepts, it involves having some (degree of) moral value, being the subject or object of a moral ought proposition (of some degree of binding force or lexical rank), or being the subject or object of a moral reason for action (of some degree of weight or lexical rank). And such a moral property is *basic* if it satisfies the following set of conditions. The property is generally *natural*—that is to say that its possession arises as a consequence of the normal course of environmental processes, such as birth, growth, or maturity. The property is generally *inalienable*—that is, it is not liable to be easily extinguished or transferred by its possessor to another, without the possessor themselves ceasing to exist. The property and its implications, generally, have *great weight* or *lexical priority*: we rarely (if ever) have reason, all things considered, to act in any way inconsistent with the value, ought proposition, or reasons for action constituting the basic moral property. Thus, collectively, these three conditions mean that when articulating a moral theory, the distribution of the property (equal or unequal) across a set of individuals will be theoretically *foundational*, or very close to foundational. This is to say that other major claims within a theory will be implied, or at least conditioned, by the fact of that distribution.

The advantages of addressing the proposition Basic Equality, rather than a particular concept(ion) of basic equality, are threefold.

First, Basic Equality is entirely agnostic with respect to the various formulations of basic equality listed above: it would be true if we are equals in any one, some, or all of the properties referred to by such formulations, or indeed any other properties not so referred, so long as they fit the criteria for being morally basic. In other words, we only have to justify that we share *one* basic

7 Kirby, "Two Concepts of Basic Equality"; Waldron, *One Another's Equals*, 3. See also Arneson, "Basic Equality," 30.

moral property equally in order to justify that we are one another's equals in *some* basic moral property—that is, Basic Equality.

Second, as I shall demonstrate in the next section, this is appropriate because the problem of basic equality is that there is a strong argument against this very possibility. It is not merely an argument against any particular concept(ion)s of basic equality, but any possible concept(ion). It is an argument against Basic Equality. This is what makes the argument, *prima facie*, so devastating.

Yet, third, despite this, the agnosticism of Basic Equality also permits us, in attempting to solve this problem, to be entirely open with respect to *both* the possible basic moral property or properties in which we might be equals *and* their possible nonnormative grounding property or properties. Any *pair* will do. Thus, to recast the challenge of solving the problem of basic equality, we are actually searching for a pair (or pairs) of properties. We are not merely searching for any possible nonnormative basis for a particular, predefined basic moral property. Indeed, and perhaps most importantly, we may be led to infer the former from the latter. We may *discover* (or at least clarify) the sense(s) in which we are one another's equals by such attempts to defend the proposition Basic Equality.

2. THE ARGUMENT AGAINST BASIC EQUALITY

The argument against Basic Equality, as the functionally defined proposition stated above, starts with the assumption that a human's possession of any basic moral property must be grounded in her possession of a nonnormative property—that is, its "basis."[8] Sometimes, however, in discussions of basic equality, authors speak of a relation of "supervenience" rather than grounding.[9] Both relations imply a modal relationship such that if fact *A* (fully) grounds fact *B*, then, necessarily, if *A*, then *B*. Yet, unlike supervenience—at least as it is typically understood today—a grounding relation is asymmetric, since it does not also follow that if *A* grounds *B*, then necessarily, if *B*, then *A*.[10] And, further, unlike "one-way supervenience," which is also asymmetric, it implies more than necessity between *A* and *B*. It implies an "explanatory" or "determinative"

8 Rawls, *A Theory of Justice*, 441.

9 Carter, "Respect and the Basis of Equality," 549; Thomas, "Equality within the Limits of Reason Alone," 540–41. Jeremy Waldron is quite adamant that he will continue to use the term "supervenience" in framing the problem of basic equality in the face of the emerging contemporary literature on grounding. However, when explaining what he means by "supervenience," he cites a definition from Simon Blackburn from 1999 that many today would take to characterize well the grounding relation in contrast to supervenience (Waldron, *One Another's Equals*, 61n30).

10 See Berker, "The Unity of Grounding," 735.

connection: "a movement, so to speak, from antecedent to consequent."[11] Thus, if *A* grounds *B*, then the fact that *A* obtains *explains why B* obtains. *B* obtains *in virtue* of (or *because* of) A.[12]

Further, discussions of basic equality may benefit from two key distinctions between types of grounding.[13] First, there is *full* and *partial* grounding: *A* fully grounds *B* if *A* by itself is sufficient to ground *B*. *A* is a partial ground for *B* if *A* and some other fact, or facts, are sufficient to ground *B*. Second, we might speak of *mediate* and *immediate* grounding: *A* mediately grounds *B* if *A* only grounds *B* by grounding *C* (that may ground *D*, . . . *E*, . . .) that grounds *B*. *A* immediately grounds *B* if no such "chaining" of grounding conditions is required.

Thus, applying these clarifications, the starting assumption in the argument against Basic Equality is that if a human has a basic moral property, then that fact must be partly grounded by the fact that she has a nonnormative property, which will partly explain why the human has the basic moral property. Such grounding will only be partial because we might assume that the fact that the human has a basic moral property will be fully grounded by the conjunction of (at least) two facts: that she possesses the nonnormative property (a nonnormative fact) *and* a general moral principle (or "bridge law" as Gideon Rosen puts it) that, for any object (or relevant subset of objects, of which the human is a member), possession of this nonnormative property entails possession of the basic moral property (a normative fact).[14] Thus, we expect a structure like:

> *Basis of a Basic Moral Property*: For any particular object *P*, *P* has basic moral property *M because* (1) *P* has nonnormative property *X*, *and* (2) for any object, the fact that it has *X* entails the fact that it has *M*.

So, for example, one might argue that the fact that a particular human being has the capacity to reason (a nonnormative property) partly explains why she

11 Fine, "Guide to Ground," 38; see also McLaughlin, "Varieties of Supervenience."

12 Those who articulate the problem of basic equality in terms of supervenience tend to introduce a further constraint on any possible solution—that is, any putative basis must not merely co-vary with the property of being a moral equal but also be "relevant" (Thomas, "Equality within the Limits of Reason Alone," 544; Carter, "Respect and the Basis of Equality," 541). But this relation of "relevance" between subvening and supervening properties would seem to be merely a search for an *explanatory* relation between them. In which case, the concept of grounding is simply being introduced by another name.

13 Fine, "Guide to Ground," 50–51.

14 Rosen, "Metaphysical Relations in Metaethics." I here remain as agnostic as possible with respect to the ongoing debate as to such a principle's modal status, explanatory role, and significance for debates between naturalism and nonnaturalism. In addition to Rosen, see Morton, "Grounding the Normative"; and Berker, "The Explanatory Ambitions of Moral Principles."

has moral worth (a plausible basic moral property). It only partly (rather than fully) explains why she has moral worth because it is also explained, putatively, by a general moral principle that any object (human or otherwise) that has the capacity to reason has moral worth. If these claims are true, then it is a short step to establishing something close to Basic Equality: assuming that not just *this* human, but all (or nearly all) humans have the capacity to reason, then they all (or nearly all) must have moral worth. The trouble, of course, is that the very last step required to establish Basic Equality does not follow: it does not follow that all (or nearly all) humans have *equal* moral worth.

This last step is the real difficulty in establishing Basic Equality. On the one hand, *prima facie* the most plausible candidate nonnormative properties for *X* within the structure Basis of a Basic Moral Property are scalar: rationality, sentience, intelligence, empathy, agency, a sense of justice, a good will. Even "conceptions of the good" can be more or less complete.[15] And, further, as a matter of empirical fact, human beings do possess such properties to varying degrees.[16] On the other hand, there is what Richard Arneson calls a "pressure of reason" to presume that if a scalar nonnormative property grounds a scalar basic moral property, then *ceteris paribus* a greater (or lesser) degree of that nonnormative property must surely proportionally ground a greater (or lesser) degree of the basic moral property.[17] As Louis Pojman puts it, "If *P* constitutes human worth, then it would seem that the more of *P* that a person has, the better he or she is. . . . If reason is really all that makes us valuable, then the more of it the better. . . . If our ability to will the good is what gives us value, then it would seem that some people are more valuable than others because they have greater ability to will the good than others."[18] This presumption is defeasible. Hence, the "*ceteris paribus*." But it does mean that the onus of proof is shifted onto the proponent of Basic Equality.

Assuming any human's possession of any basic moral property is grounded in such a scalar nonnormative property but that humans tend to hold the latter to different degrees, it follows that *ceteris paribus*:

> *No Basic Equality*: All (or nearly all) humans are not one another's equals in their possession of *any* basic moral property.

15 Schaar, "Some Ways of Thinking about Equality," 867.

16 Spiegelberg, "A Defense of Human Equality," 106.

17 Arneson, "Basic Equality," 36.

18 Pojman, "A Critique of Contemporary Egalitarianism," 484–85. See also, Carter, "Respect and the Basis of Equality," 541; Christiano, "Rationality, Equal Status, and Egalitarianism," 56.

3. "NONSCALAR" STRATEGY: RANGE PROPERTIES AND BARE PROPERTIES

A number of replies to the argument against Basic Equality have been proposed. One might reject the starting assumption that the possession of a morally basic property needs to be grounded in any nonnormative fact about humans: it is just an ungrounded fact.[19] Or insofar as it is grounded, it is grounded in our "humanity" and whatever then grounds the fact that we are human.[20] Or one might argue that our equal possession of a particular morally basic property is not even a fact about humans, but rather it is a proposition that we each (should?) choose to assume about humans when engaging with them.[21] Or one might accept the starting assumption of the argument above but argue that the nonnormative property may be theological or transcendental, thus voiding any presumption that it is likely to be spread unevenly among us.[22] Or one might argue that it follows as a formal principle of rule application.[23] Or one might look to universal prohibition against treating others as inferiors.[24] Or one might even concede that Basic Equality may be false but hold that some (many?) moral arguments need not actually rely on it to justify many forms of so-called egalitarian rights.[25]

I shall not rehearse these arguments here. I take them all to meet convincing counters already supplied in the literature (cited in the respective footnotes above). Instead, I shall focus my critique on what is currently considered by many to be the most promising line of thought—that is, the nonscalar strategy. The nonscalar strategy rejects the claim that only scalar properties are plausible bases of basic moral properties, and aims to identify a nonscalar property instead. If the property lacks scalarity, so the thought goes, Arneson's pressure

19 Gosepath, "On the (Re)Construction and Basic Concepts of the Morality of Equal Respect," 125. But see Husi, "Why We (Almost Certainly) Are Not Moral Equals," 388.

20 Vlastos, "Justice and Equality"; Frankena, "The Concept of Social Justice." But see Wilson, *Equality*, 93; Schaar, "Some Ways of Thinking about Equality," 875–82.

21 Macdonald, "Natural Rights"; Arendt, *On Revolution*. But see Waldron, *One Another's Equals*, 55–61.

22 On the theological, see Waldron, *One Another's Equals*, 175–214. But see Thomas, "Equality within the Limits of Reason Alone," 539–40. On the transcendental, see Kant, *Groundwork of the Metaphysics of Morals*. But see Williams, *In the Beginning Was the Deed*.

23 Westen, *Speaking of Equality*; Lucas, "Against Equality"; Frankfurt, *On Inequality*. But see Waldron, *One Another's Equals*, 66–83.

24 Sangiovanni, *Humanity without Dignity*. But see Floris, "Two Concerns about the Rejection of Social Cruelty as the Basis of Moral Equality."

25 Husi, "Why We (Almost Certainly) Are Not Moral Equals," 381–84; Steinhoff, "Against Equal Respect and Concern, Equal Rights, and Egalitarian Impartiality." But see Waldron, *One Another's Equals*, ch. 1.

of reason will not so much disappear as work in reverse.[26] If a nonscalar nonnormative property grounds a morally basic property, then *ceteris paribus* having the same nonscalar nonnormative property must surely ground the same (or same level of) the basic moral property. The initial challenge, however, is to find a relevant nonnormative property, given that *prime facie* most of the plausible bases of a basic moral property are scalar.

In response, the literature actually bears out two different types of candidate nonscalar properties, but they are currently not distinguished. The first property is what John Rawls calls "a range property."[27] If we can define a range of degrees in a scalar property, then there is a property of falling inside that range. This latter property is nonscalar. It is a range property. For example, take the property of having length: it is scalar. Things can be longer or shorter. However, if we simply define a range of length values (e.g., more than ten meters), then objects will either fall within this range (by having any length more than ten meters) or not. This means that objects can have nonscalar properties in virtue of having some particular degree of a scalar property within a defined range. More formally:

> *Basis of a Range Property* R: For a particular object *P* and particular range of degrees *N*, *P* has the nonscalar property *R* of having scalar property *G* within range of degrees *N because* (1) *P* has scalar property *G* to a particular degree *D and* (2) particular degree *D* falls within the range of degrees *N*.

A range property, however, should be disambiguated from what I shall call a *bare property*. Instantiations of a scalar property may always come in degrees, but we can define a nonscalar property that is simply having that property to *any* degree. For example, the property of having length is scalar, but the property of having *a* length is nonscalar. One either has *a* length or not. The key difference between a range property and bare property is that the latter does not require us to define a threshold, or in fact any range at all. More formally:

> *Basis of a Bare Property* B: For a particular object *P*, *P* has the nonscalar property *B* of having scalar property *G* to some degree *because P* has scalar property *G* to any particular degree *D*.

Now, with these two different nonscalar properties in hand, and drawing on the distinction flagged above between mediate and immediate grounding,

26 Williams, *In the Beginning Was the Deed*: "For differences in the way people are treated, some general reason should be given" (98).

27 Rawls, *A Theory of Justice*, 444.

we can distinguish within the current literature two different (sub)strategies to ground our basic equality.

One strategy aims to ground our possession of a basic moral property immediately on our possession of a range property. The other strategy aims to ground our possession of a basic moral property immediately on our possession of a bare property. It is true that our possession of this latter bare property is itself likely to be based on our possession of a range property. This leads to the common conflation of the two properties.[28] But, strictly speaking, in the latter case, there is a *grounding chain* from range property *to* bare property *to* basic moral property.

Rawls adopts the first strategy. He states:

> The question of equality arises. The natural answer seems to be that it is precisely the moral persons who are entitled to equal justice. Moral persons are distinguished by two features: first they are capable of having (and are assumed to have) a conception of the good (as expressed by a rational plan of life); and second they are capable of having (and are assumed to acquire) a sense of justice, a normally effective desire to apply and to act upon the principles of justice, at least to a certain minimum degree.

Now at first glance, one might be tempted to see Rawls proposing something like the second strategy with a grounding chain: a range property (having a capacity to have a conception of the good and sense of justice, at or above a certain minimum degree) grounds an intermediate property (moral personality), and having such moral personality grounds our equality. However, Rawls offers no definition of the term "moral personality" beyond its two distinguishing features. It is not a new property grounded by the range property. Instead, "moral personality" *is* his elliptical name for the range property.[29] Hence, we should be able to eliminate reference to moral personality, effectively taking him to claim that "it is precisely [that those *capable of having a conception of the good and sense of justice, at least to a certain minimum degree*] are entitled to equal justice."

28 See Waldron, *One Another's Equals*, 118–19. Waldron appears to define "range property" as a bare property that is *grounded by* what I have termed a range property.

29 Of course, I could be wrong, interpretatively, about Rawls. Moral personality may be a new property that we have *in virtue* of having the range property of meeting the minimum threshold of relevant capacities. But if moral personality is not the range property itself (as I define "range property"), but instead a different property that we have in virtue of having that range property, then without any further elaboration, Rawls has given us no reason to hold that this new property is also nonscalar. Moral personality (still yet to be defined) might well come in degrees. Of course, Rawls may well then hold that he is concerned with the *bare* property of having a moral personality to some (any) degree. However, he is then just as vulnerable to the critique below (section 4 below).

Eliminating "moral personality" as a term, however, makes the problem of any such immediate grounding strategy transparent. This strategy cannot explain why being within the relevant range (having a capacity to have a conception of the good and sense of justice to certain minimum degree) is special or, at least, special enough to ground a morally basic property that one did not necessarily have before. For any range *N* that supposedly grounds our basic morally equality, I can propose another proximate range $N+1$, or $N-1$, and ask what explains the difference in treatment between N and these ranges. However, since the grounding is *immediate*, then *ex hypothesi*, there is no other fact to offer an explanation. Without an explanation, the immediate grounding on the range property seems entirely arbitrary.[30]

Contemporary approaches, therefore, learn this lesson from Rawls's failure and actually adopt the second strategy—even if only implicitly. For example, take Thomas Christiano's argument. Christiano begins by identifying a scalar feature that he takes only humans to have to *any* degree (i.e., being a "rational being"): "Rational beings are capable of reflection on the norms that govern their behavior and the norms that govern the formation of belief and inference . . . a higher order capacity that does not seem present in the case of other higher mammals."[31] Christiano argues that being such a rational being is a "discontinuity" with other animals. By this, he seems to imply that there is another scalar property (a continuity) that we do share with other animals, although to a higher degree (e.g., other reasoning capacities), but at a certain threshold of these capacities, a new property (i.e., being capable of reflection on norms) is triggered. While this latter property is triggered by being above this threshold (i.e., the range property of being above that threshold of other reasoning capacities), it is *not* that range property. It is a different, *bare* property of having *some* (i.e., *any*) capacity to reflect on norms.

George Sher's argument has the same structure. Following Bernard Williams, he identifies consciousness or "subjectivity" as a basis for a morally basic property in which we are equals. Of course, this is a scalar property: people are more or less conscious of themselves and the world around them. However, possessing this scalar property to different degrees still entails that we have a bare property—that is, being a subject *with* (any) mental contents: "If the reason we are moral equals is simply that each of us has (is?) a subjectivity of a certain sort . . . then any variations in the contents of our beliefs and aims, *and in the capacities that gave rise to these*, will simply drop out as irrelevant."[32] The

30 Arneson, "What (If Anything) Renders All Human Persons Morally Equal?," 108–9.

31 Christiano, "Rationality, Equal Status, and Egalitarianism," 62.

32 Sher, *Me, You, Us*, 36–37.

latter italicized phrase is referring once again to the presumed scalar property (*qua* "capacities") that we must have within a certain range to have the bare property (i.e., subjectivity). But once again, that range property should not be confused with the bare property it, in turn, grounds.

Tom Parr and Adam Slavny argue that our possession of the "capacity for a conception of the good" (CCG) grounds our morally basic property. One has a CCG "if and only if she can form, revise, and pursue beliefs about the good on the basis of critical deliberation."[33] Now Parr and Slavny expressly contrast scalar aspects of this CCG (e.g., exercising it well) and the nonscalar property of exercising it "*tout court*": "Our interest in the *mere* exercise of a capacity does not vary according to how well we exercise it, as exercising a capacity poorly entails exercising it *tout court* just as [much] as exercising it well."[34] CCG, therefore, appears to be a bare property. However, they then say, "The capacity to pursue a conception of the good *tout court* is a range property. There is a threshold below which an individual lacks the necessary subvenient properties for the CCG."[35] But by their own definition, CCG does not identify a threshold on any scalar property. It is the property of having *any* capacity for a conception of the good, not of having that capacity above a threshold. Instead, a better view is that, once again, having this *bare property* is grounded by some range property but is not itself that range property.

Jeremy Waldron might also profit from distinguishing between range and bare properties. His recent gloss on Rawls's definition is as follows:

> Rawls's idea involves a relationship between two associated properties. There is the property *R*, which operates in a binary way (either you have *R* or you don't), and property *S*, which is a scalar property admitting of differences of degree. We say that *R* is a range property with respect to *S*, if *R* applies to individual items in virtue of their being within a certain range on the scale indicated by *S*. In the simplest cases, *R* is like a threshold. If you are over a specified threshold on scale *S*, you qualify for property *R*. But the range may have an upper limit as well, or it may be configured in a more complicated way in a two- or *n*-dimensional model.[36]

Waldron here appears to define "range property" as a property *R* that one possesses "in virtue of" possessing a scalar property *S* at some degree that falls inside a specified range of degrees of *S* (e.g., above a "specified threshold" and

33 Parr and Slavny, "Rescuing Basic Equality," 842.

34 Parr and Slavny, "Rescuing Basic Equality," 843.

35 Parr and Slavny, "Rescuing Basic Equality," 843–44.

36 Waldron, *One Another's Equals*, 118–19.

below "an upper limit"). By contrast, adapting Waldron's variables, I simply define "range property" as the property *R* of possessing scalar property *S* at some degree that falls inside such a specified range of degrees of *S*. In other words, Waldron's range property is any property that is *grounded* (qualified for) by possessing what I term to be a range property.

In itself, of course, such a difference of terminology should not be a problem. However, Waldron's definition of "range property" does not entail that a range property is nonscalar ("binary"). For example, just because I pass the threshold level of cognitive capacities (*S*) to have moral agency (*R*), does entail that such moral agency will be nonscalar. Indeed, we might expect such moral agency to come in degrees, in part, as a very function of further higher degrees of cognitive capacities above the relevant threshold. So if being "nonscalar" is meant to be a necessary feature of a range property, as Waldron also claims, then his definition fails. But of course, there is a nonscalar property that we do necessarily possess when we possess a so-termed range property, that is, the bare property of having *any* degree of that so-termed range property. And, indeed, upon illustration, that is, precisely what Waldron is interested in. Thus, according to Waldron, Hobbes's explanation of basic equality involves a scalar property "strength of body" and a so-termed range property, *but actually bare property*, "for each person *P*, the property someone else has of being a non-dismissible mortal threat to *P*"—that is, being some (any) degree of nondismissible mortal threat to *P* (after all, "non-dismissible" just means that we have reason to pay some [any] degree of attention to the threat, but of course, some threats may deserve more attention than others).[37] Or for Locke, according to Waldron, the scalar property is reason in general and the so-termed range property, *but actually bare property*, is "the ability to know God through engaging in abstract thought"—that is, having some (any) ability to know God through engaging in abstract thought.[38]

Finally, the best-known contemporary attempt to articulate a basis of basic equality could also benefit from a distinction between a range property and bare property. Ian Carter states:

> My suggestion, then, is that equality of certain entitlements is justified because those entitlements should be assigned on the basis of personhood, and while the agential capacities on which the ascription of personhood is based are themselves ultimately scalar properties (as they must be, on any naturalized account of the basis of Kantian respect), it is appropriate to treat personhood as a range property because it is

37 Waldron, *One Another's Equals*, 120–21.

38 Waldron, *One Another's Equals*, 121–22.

> appropriate to show opacity respect toward beings that meet a certain absolute standard of moral agency.[39]

Here, like Rawls, Carter first refers to "personhood" as if it might be a bare property that is grounded by ("based" on) having a different range property (i.e., being within a threshold of scalar agential capacities). However, he clarifies that, once again like Rawls, he is just naming the latter range property "personhood." Thus, it seems, he is arguing that this range property immediately grounds the morally basic property or properties in which we are equals. But this, then, would leave Carter in the same position as Rawls. Without an intermediating fact to explain why a *particular* threshold of agential capacities triggers such morally basic properties, this immediate grounding relation is arbitrary. It does not help his argument that one of the morally basic properties that also might be triggered by reaching this threshold is a right to "opacity respect," if that threshold itself remains unexplained.[40]

However, Carter *does* have a possible explanation. He just does not make it explicit. Later in the paper, he argues that what grounds our duty of opacity respect is not "being within some threshold of agential capacities" *qua* range property. It is instead "agency itself" *qua*—at least as I suggest—bare property.[41] What is this bare property? We have to infer the meaning from Carter's reasoning, but it seems to be whatever nonnormative capacity only humans have, which explains why only humans are liable to some (any) reactive attitudes such as praise, blame, and resentment. The duty of opacity respect is then needed precisely so that these attitudes are not "dismantled" or "explained away." Thus, on my reading, Carter—unlike Rawls—has an (extended) chained version of the second strategy, where our "being within some threshold of agential capacities" (*qua* range property) grounds some new nonnormative property "agency itself" (*qua* bare property), which in turn grounds both the moral property of

39 Carter, "Respect and the Basis of Equality," 554.

40 Alternatively, Carter can be read as arguing that the right to opacity respect not merely explains why we cannot assess differences above the threshold but also identifies the threshold: in other words, we know someone has met the threshold of empirical agential capacities when they have this right to opacity. But when we go to then search for the nonnormative property that grounds the opacity, it ends up being the meeting of this very undefined threshold. This is because Carter argues that the imperative of opacity respect is grounded on the need for outward dignity (another moral property), which is grounded by "dignity as agential capacity" (another moral property). What is dignity as agential capacity? It is the dignity we have "in virtue of our agential capacities." However, what agential capacities ground such dignity? An individual possesses "*dignity as agential capacity* [when they] possess at least a certain absolute minimum of the relevant empirical capacities" ("Respect and the Basis of Equality," 556).

41 Carter, "Respect and the Basis of Equality," 558.

being liable to reactive attitudes (another bare, but this time moral, property) and also the right to opacity respect (*qua* another bare moral property), and together these latter two moral properties ultimately ground "equality of certain entitlements" (*qua* equality in a morally basic scalar property).

4. BARE PROPERTIES: A FALSE HOPE

So my interpretive claim is that most current purported proponents of the *range property only* strategy must be deploying a *chained bare property* strategy instead. The *prima facie* attraction of such a strategy is that it solves the arbitrariness problem of the range property only strategy, and thus eliminates the pressure of reason to accord proportionate significance to the variations in the scalar property that underlies the range property. The relevant threshold marks a nonarbitrary, nonscalar distinction with respect to that latter scalar property: below this threshold, an object does not have the bare property; above this threshold, it does. So, for example, following Christiano: below a threshold level of particular rationality, rationality is insufficient to ground any capacity to reflect; above that threshold, rationality is sufficient to ground at least *some* capacity to reflect. This is, indeed, a nonarbitrary, nonscalar distinction between two different ranges of rationality.

The problem for these proponents of the chained bare property approach, however, is that instead of solving the pressure of reason problem, they have merely *shifted* it. While they have explained why we can safely ignore differences in the particular scalar property that underpins the range property above the threshold (e.g., for Christiano, rationality), they have only done so by introducing a *new* scalar property that underpins the bare property (e.g., for Christiano, the capacity to reflect). Their approach simply moves the pressure of reason problem. They now must explain why differences in that new scalar property do not ground proportionate differences in the moral property.

To continue with Christiano, he assumes, following the "late scholastics," that "inasmuch as human beings are rational beings [that is, capable of reflection], . . . persons are not made merely for each other's use. The idea here is that each person has a kind of original right against others." He then simply jumps to a claim that this right will be equal for each human: "It does not admit of the idea that one may treat one person some of the time or in some respects as a means while others may never be treated as mere means."[42] But why? It is perfectly possible to think of a hierarchy of human instrumentalization; in fact, arguably, that has been the dominant political theory in history. So, *if* being a

42 Christiano, "Rationality, Equal Status, and Egalitarianism," 63–64.

rational being is so important as to ground a right not to be instrumentalized, then why does greater rationality (*qua* capacity to reflect) not ground a greater right not to be instrumentalized? We could still conclude that all (or nearly all) humans have a right not to be instrumentalized, but this just means we have the bare property of having such a right of *some* lexical rank or scope, but not necessarily of *equal* rank or scope.

Similarly, George Sher argues that the bare nonnormative property of having a "consciousness" grounds certain interests "in (say) accomplishing his rational ends, or in having the opportunities or resources to do so." He then asserts that we have equality in some scalar morally basic property: "As long as two people both meet this requirement [having a consciousness], the fact that their plans differ in complexity and sophistication will not mean that one has more of an interest in succeeding than the other."[43] But once again, why? Sher never explains. So, the pressure of reason problem reemerges: If consciousness is so important, then why is more consciousness not more important?

Parr and Slavny claim that having "CCG *tout court*" grounds having "a weighty interest in being the author of their own lives." They then outsource the key question of the comparative equal weight of that interest to theories of "self-authorship": "We will not develop a specific account of our interest in self-authorship here. . . . Being the author of one's own life, on most plausible conceptions, is an interest in exercising a capacity *tout court*, rather than exercising it well."[44] But once again, this will not do since in debates about Basic Equality we are calling into question this very intuition that such conceptions tend to take for granted. We have to answer the riposte: If having a CCG grounds an interest in self-authorship, then why does having a greater (more complete, more internally consistent, more accurate, more reflective and self-originating) CCG not ground a greater interest in self-authorship?

Much the same may be said for Waldron's own illustrative examples, although I take him, himself, to ultimately adopt a theological strategy to defending Basic Equality.[45]

But what of Carter's approach? Carter's argument—at least as I interpret it—is different from the others. Carter does not try to argue that the possession of a nonnormative bare property *immediately grounds* the possession of equal degrees of some morally basic scalar property in each of us. Instead, he argues that possessing a nonnormative bare property ("agency itself") immediately

43 Sher, *Me, You, Us*, 40.

44 Parr and Slavny, "Rescuing Basic Equality," 844.

45 Waldron, *One Another's Equals*, 185.

grounds possession of a corresponding moral bare property ("liability to reactive attitudes"):

> Possession of nonnormative bare property *A grounds* possession of moral bare property *M*.

Carter, then, is delaying the further step—where possession of a bare property grounds possession of equality in some scalar property—until it is between types of moral property:

> Possession of a moral bare property *M grounds* possession of equality in some scalar moral property *X*.

In fact, by delaying this move until it is between moral properties, he is able to argue that such equality in a scalar moral property is *grounded by the combined effect of two bare moral properties*, both of which are themselves grounded by the nonnormative bare property. He argues that, on the one hand, agency itself grounds some (any) liability to reactive attitudes. On the other hand, it also grounds the imperative of opacity respect. Such opacity respect is an "external perspective," "evaluative abstinence," a "blindness … that avoids evaluation of the agential capacities on which moral personality supervenes."[46] Without opacity respect, agency itself is at risk of being dismantled (and, by assumption, we have reason not to dismantle agency). However, such opacity respect neutralizes the moral significance of any degrees of difference in one's liability to reactive attitudes. Thus, the *combined effect* of these two bare moral properties, both grounded by the same nonnormative bare property (agency itself), is that we only have reason to treat people as if they are liable to reactive attitudes to the same (full) extent. Collectively, therefore, they ground equality in a further moral property: equality in entitlements. So, in the end, we have:

1. Possession of nonnormative bare property *A grounds both* possession of moral bare property *N and* possession of moral bare property *M*.
2. *Both* possession of a moral bare property *N and* possession of moral bare property *M ground* possession of equality in some scalar moral property *X*.

There is something promising about the prospect that our final equality in a moral property might be grounded by some relation between two or more other moral properties, which are themselves grounded by a nonnormative property. However, Carter's argument as it stands has two flaws.

46 Carter, "Respect and the Basis of Equality," 552.

The first is that, at best, Carter's argument establishes that any individual with agency is due some (any) degree of opacity respect by others (a bare property). However, it is not clear exactly why each individual is due the *same*, equal degree of opacity respect. After all, the imperative to treat another with opacity respect may come in degrees too. It may be true that *adherence* to the imperative does not come in degrees—one either treats another with opacity respect or not—but the *strength or lexical priority* of that imperative might vary across individuals. Thus, it might be *more* imperative to treat one individual with opacity respect, despite the costs to them or others, than another individual. And indeed, if what explains the imperative in the first place is the value of agency—giving rise to the reason not to dismantle that agency—then surely there is a stronger, more stringent imperative to protect the person with *more* agency since dismantling it would be a greater loss than dismantling *lesser* agency.

However, even if Carter can respond to this first problem, he will still be left with a second problem that goes to the argument for needing opacity respect at all: "looking inside" at people's degrees of agency—that is, the underlying nonnormative property that grounds liability to reactive attitudes—will not necessarily lead to such dismantling of agency itself.[47] It is perfectly possible to look inside people's agency and distinguish between those aspects of their lives that ought to ground reactive attitudes and those that ought not. Take, for instance, the common law mitigatory defense of "provocation," which converts an act that is otherwise murder to manslaughter because of some "sudden or temporary loss" of self-control.[48] Once provocation is established, however, the law does not consider the individual's agency to be dismantled or explained away. The individual is still held responsible for their act. Thus, they are still liable to be convicted for manslaughter. But they are thought to be less culpable than if they had acted in a premeditated or purely malicious fashion. This reasoning might entail a kind of basic inequality—that is, we are unequally liable to reactive attitudes. But the point is that this conclusion has not come at the putatively unacceptable cost of dismantling agency since agency remains—although it has been reduced or qualified.[49]

47 A deeper critique would simply press that in relying upon Strawson, Carter inherits his problem: the undesirability of such dismantling does not mean it is unwarranted. See Wolf, "The Importance of Free Will."

48 R v. Duffy, [1949] 1 AER 932 (CA), per Devlin J.

49 A similar point is made by Arneson, "Basic Equality," 48.

Christopher Bennett has recently sought to defend Carter against a similar objection.[50] As Bennett formulates the objection, the imperative of opacity respect appears to be inconsistent with defenses such as duress, loss of control, and automatism, as well as court practices of taking into consideration agential factors in pretrial, sentencing, and parole hearings.[51] Bennett takes this to be an unattractive, counterintuitive implication for Carter. Coming to his defense, however, Bennett argues that contrary to the objection, such court practices are consistent with adhering to the imperative of opacity respect. He argues that this is so *because they are consistent with respect for the party's agency*. They are consistent with respect for the party's agency because they hold the party "accountable" only for what they have truly "done," in the sense of what they are properly "answerable for" as a function of their "intentions" and "reasons" for acting. In particular, excusatory defenses (like provocation) simply determine "whether in a full sense one can be said to have done the thing in question."[52]

The problem for Bennett (and thus also Carter) is that one can concede that such court practices are indeed consistent with respect for the party's agency, but press that they are so precisely because they are inconsistent with opacity respect—that is, they involve looking inside and assessing the *degree* of agency that a person is exercising in the relevant scenario, and thus what they are accountable for and in what way. The burden of Carter's argument, by contrast, at least as I understand it, is to show that any such attempt to assess the degree of an individual's agency will entail a dismantling of their agency altogether. Indeed, one might be tempted to turn Bennett's argument around on Carter: since such court practices are not merely consistent with but *necessary* for respect of the relevant party's agency, and yet they are inconsistent with the imperative of opacity respect, it follows that in these circumstances opacity respect is inconsistent with respect for the relevant party's agency.[53]

50 Bennett, "Intrusive Intervention and Opacity Respect." I thank an anonymous reviewer for bringing this to my attention.

51 Bennett, "Intrusive Intervention and Opacity Respect," 270–71.

52 Bennett, "Intrusive Intervention and Opacity Respect," 271.

53 At the end of his argument, Bennett does assert that despite all the explicit assessment that goes on in such court practices of a party's degree of control over themselves and their knowledge, rational functions, and other capacities, an underlying level of opacity respect remains because the practices "involve . . . taking the exercise of one's agency at face value, not second-guessing or pre-empting it. . . . They are compatible with the idea that a person is defined by how they will to present themselves, and that 'the mess inside' that issues in such action should be treated as opaque" ("Intrusive Intervention and Opacity Respect," 271). However, it is difficult to assess this parting assertion since Bennett does

5. A REVIVED STRATEGY: RELATIVE POTENTIAL

With the failure of both the bare and range property strategies, and indeed the current pessimism about other approaches, are there any grounds for optimism in solving the problem of basic equality?[54]

The chained bare property strategy was an improvement over the range property only strategy because it suggested that progress can be made by focusing on a nonnormative scalar property that begins in the human range—that is, all (or almost all) human beings have at least some (any) degree of this nonnormative scalar property—which thus gives us a nonarbitrary range-defining threshold for some further, deeper, nonnormative scalar property possibly shared with other beings. The strategy failed, however, because it could not explain why we would then ground any basic moral property simply on the *bare* fact of having some (any) of that nonnormative scalar property, rather than on the degrees of that property. It could not escape the pressure of reason problem once shifted, not even with opacity respect.

We might conclude, therefore, that we need an approach that does not ignore such differences of degrees, but rather in some way works with such differences to ground an equality.[55] But how might this be possible?

My suggestion is the idea of *relative* potential. Let me illustrate the concept in the abstract first, and then I will deploy it in the context of basic equality. Imagine a line of boxes of different sizes from, let us say, one cubic meter up to one hundred cubic meters. Each box, therefore, varies in the scalar property of volume. Now we might begin to fill each box with contents. Each box clearly has *different* storage potential in *absolute* terms: each can hold a different volume of contents measured by cubic meters. However, each box has the *same* potential in *relative* terms: no matter their absolute volume, each and every one has the potential to be filled 1 percent or 10 percent or 50 percent, all the way up to 100 percent. Hence, in relative terms, we can make best use of a box (100 percent volume) or worst use (0 percent) of a box; it can be equally full or equally empty, regardless of its absolute volume. This is a kind of equality: each box has the same potential to be filled to any *relative* extent.

---

not explain what exactly remains taken at "face value," or what part of the "mess" is not open for court assessment.

54 On other approaches, see notes 14–20 above.

55 Waldron also gestures in this direction with his discussion of scintillation, but he only demonstrates how current secular conceptions of basic equality appear to do this, rather than aiming to explain *why* (*One Another's Equals*, ch. 4). His primary argument remains theological.

This is quite an odd example, I concede. We are not boxes. But consider this somewhat neglected passage, one of the first contemporary explorations of the problem of basic equality, by Herbert Spiegelberg.[56] He states:

> It should, however, not be overlooked that among the moral values there are some which involve a *potential equality* in one important respect. If we exert ourselves for a certain cause with all the energy at our disposal, however weak it may be, the outcome of such exertion will certainly vary. But the intrinsic ethical value of our effort, as distinguished from the value of the result, will not depend upon the latter. . . . The moral value of our effort, then, depends exclusively upon the question how much of our momentary intellectual and moral energies was used in the attempt to ascertain and to realize the right goal. The absolute amount of our energies and of our effort is immaterial. It is only the relation between them which counts. Now these effort-values reflect also upon the agent. It is this fact which gives every agent equal access to the moral values consequent on moral effort. In the court of this particular value he faces no handicaps. Everybody who is able to run at all is given an equal chance. The tasks assigned to different individuals may be very different. In fact, the higher the abilities, the more exacting will be the demands; the smaller the means the more lenient will be the expectations. All that matters is: how big were our efforts in proportion to our unequal and varying momentary equipment?[57]

I take the underlying logic of Spiegelberg's argument to be the same as our somewhat odd box example above. In virtue of having some degree of agency, all (or almost all) human beings, unlike other animals, will have the potential to perform actions of ethical value. All of us can achieve a range of ethical values given our agential ability. However, this range is liable to be different for different human beings. For some (small boxes), their potential is small: no matter how such individuals use their agential ability, they can only achieve goals of minor ethical value at best. However, for others (large boxes), their potential is

56 This neglect is in part because Spiegelberg himself sets this argument to one side in settling on his final argument for basic equality, where he argues that our basic equality lies in being given "an equal start" at the beginning of our lives in the overall challenge of acquiring ethical value. This is because each individual's "ethical score" will always begin at zero (116). Yet this argument has far fewer prospects of success than the one contained in the neglected passage: not only would other animals also share an ethical score of zero, but it is also hard to see how such an ethical score can ground any positive claims. Being equally nondeserving is unlikely to ground a foundational ethical claim to equal rights, respect, consideration, or concern.

57 Spiegelberg, "A Defense of Human Equality," 108 (emphasis added).

great: these individuals have the ability to attain goals of very great ethical value at best. However, just as with filling our boxes, each of us will have the same potential relative to these constraints. After all, each of us, no matter our degree of ability, can do our best with that degree of ability: achieving the best ethical goal available to us. And indeed, each of us, no matter our degree of ability, can also do our worst: achieving nothing valuable at all.[58] And each of us can act in between.[59] So we each have the same equal relative potential—that is, the ability (at least) to attain the highest ethical value available to us (or the lowest, or in between). Spiegelberg's key claim, then, *is that it is just such performance relative to potential* that really determines the moral value of our efforts. Here, he is introducing a distinction between "ethical value" *simpliciter* (or the "value of the result") and "moral value" (or "intrinsic ethical value"). He claims that our degrees of relative performance in attaining ethical value now ground degrees in the further absolute value, moral value. One's best possible performance will have the same moral value as anyone else's, regardless of any differences in absolute terms. Conversely, one's worst possible performance will have the same moral value as anyone else's, and *mutato mutandis* in proportion, for every possible performance in between. And thus, to repeat: "It is this fact which gives every agent *equal access* to the moral values consequent on moral effort. In the court of this particular value he faces no handicaps."[60]

One might concede at this point that all individuals who possess agency to *any* degree thereby have "equal ethical potential" *qua* the equal ability to attain moral value. However, one might wonder whether such equal ethical potential is a *basic moral property* in the functional sense we have defined above (section 1). One might grant that it is plausibly inalienable and natural, in the rough senses defined above, but ask what implications exactly it is meant to have. In particular, drawing on our functional test for a basic moral property detailed above (section 1), does it have implications of such great weight or lexical priority that its possession by all (or almost all) human beings is theoretically foundational?[61]

58 Or indeed, achieving the lowest degree of ethical (dis)value available, if one permits both negative and positive values.

59 For Spiegelberg's argument to give us a perfect equality, there must be a continuous range of options within the range. This is, no doubt, a requirement that grounds an objection that must be met in defending the strategy beyond this paper.

60 Spiegelberg, "A Defense of Human Equality," 108 (emphasis added).

61 Thank you to an anonymous reviewer for pushing me on this point.

Spiegelberg's own argument, I admit, is not entirely clear.[62] However, a plausible argument, and at least for the instant purposes of illustration, is the "fittingness" of ensuring that those with the *equal ability to attain moral value* have, as far as possible, *equal ability to attain ethical value*. It is true by definition that each individual will have the equal ability to achieve moral value regardless of their ability to achieve ethical value—that is, regardless of their differing endowments *qua* different levels of energy at their disposal or handicaps. However, this does not ground a reason to ignore such differences in our endowments if we can ameliorate them. Instead, quite the opposite, so the argument might go. It is only fitting that each agent's moral performance has the same *weight* in the world. Only in this way does one agent's moral performance *matter* as much as anyone else's performance.

To illustrate, one might imagine two moral twins: two individuals who perform equally well, indeed at a high level, in the domain of moral value during their lives—displaying virtues of generosity, courage, conviction, and so on. However, they do so relative to different endowments—agential and otherwise. Hence, they finish their lives having created the same degree of moral value, but very different degrees of ethical value. This is to say that their lives have had very different impacts on the world. The first twin, we might think, was born with great agential capacities and rose to become a leader in their nation, saving it from crises and steering it forward. The fact that they provided a good moral performance, given these opportunities, mattered not only in itself but also for everyone else in their nation. Their life was a "great life" (i.e., one of great importance). By contrast, the second twin, we might think, was born without such agential capacities and stayed in the village, raising a family, relating with friends, and in general just being a good person. In doing so, their good moral performance mattered somewhat—it was valued by those around them—but it certainly did not matter on the same scale as the first twin's. The second twin's life, and best efforts at being a good person, simply did not matter that much, at least compared to their moral twin. It was merely a "little life" (i.e., one of little importance)—not because they chose a different, less important path, but because that path to importance was simply not available to them, due to differences in their endowment. So the tentative suggestion is that, *ceteris paribus*, this is a kind of injustice: it is unjust that individuals equally capable of living lives of moral value matter unequally in the domain of ethical value. One might take this as a kind of very abstract argument for policies that aim to equalize our opportunities—not so much opportunities for power, welfare, resources,

62 Indeed, he seems to walk away from it, and he turns to a somewhat different argument later on that we should all have equal opportunities because we all start from an ethical score of zero (Spiegelberg, "A Defense of Human Equality," 109, 116).

or personal gain, although somewhat equal distribution of these might follow as a further consequence, but instead opportunities to "matter," to make the world a better place (or indeed fail to do so).

There are, of course, a number of possible objections to Spiegelberg's overall argument, at least so reconstructed.[63] For example, it relies on an assumption of a continuous range of options within any range to establish the equal ability to have *any* level of the further moral value.[64] It also needs to supply an account of moral responsibility *qua* "achievement" that survives the contemporary travails of compatibilism and Frankfurt-style examples.[65] One might question whether the moral implication that I have inferred does indeed follow, or if it does follow, whether it is sufficiently fundamental.[66] These are points for further research. However, my current claim is, merely, that such objections reflect progress beyond the pressure of reason problem altogether. These are not so much problems about how a scalar property could *possibly* ground an equality in another property but rather problems about whether we *actually* have a particular pair

63 Let us allay one concern, however. By "equal relative potential," Spiegelberg means "potential" in the sense of "ability" (literally "potential," as in *potentia, -ae*; "power, ability, force"). This is the same sense in which many other authors offer rational capacity, for example, as the basis of basic equality—that is, the *ability* to act rationally. For Spiegelberg it is the *ability* to achieve moral value as he defines it. To have such a kind of potential *qua* ability across a lifetime is just the spatiotemporal sum of one's ability as it varies from time to time across that lifetime. There is, to my knowledge, nothing particularly problematic in itself about positing a potential in this sense as the basis of basic equality. This is in contrast, however, to another sense of the term "potential," where "potential" means the current property that a being (e.g., a fetus) has of possibly having (or likely having) a future property (e.g., rational capacity, or indeed the ability to achieve moral value) because of both the being's current internal (or "essential") properties (e.g., genetic code) and relevant external conditions (e.g., normal, natural development inputs). This is a far more problematic use of the term "potential" since clarifying what counts as internal and relevant external properties is hard. After all, given sufficient genetic intervention (as an external condition), the fetus of *any* animal has the potential for rational capacity. However, to be clear, this problem does not apply to Spiegelberg or any account that merely uses "potential" in the current "ability" sense. For further discussion, see McMahan, "Challenges to Human Equality"; Arneson, "What (If Anything) Renders All Human Persons Morally Equal?"; Vallentyne, "Of Mice and Men."

64 Although even if this did not hold, each individual would still have equal ability to attain their highest and lowest values of that moral value.

65 Frankfurt, "Alternate Possibilities and Moral Responsibility."

66 An alternative or even complementary implication would be to ground a theory of proportional desert on the underlying moral equality, arguing that it is only fair that we are rewarded (or indeed punished) in proportion to our moral performance because we have all had equal ability with respect to moral value in our lives: "I'll say it again—it is easier for a camel to go through the eye of a needle than for a rich person to enter the Kingdom of God!" (Matthew 19:24).

of such properties, and if so, whether the latter is sufficiently morally basic. In short, my argument in this paper has merely been to justify further exploration of this style of argument from relative potential, even if one rejects the particular Spiegelbergian version. To characterize that style more formally:

1. All (or nearly all) humans possess nonnormative scalar property *X* to some (any) degree naturally and inalienably (e.g., some [any] degree of agency as a bare property).[67]
2. Possession of nonnormative scalar property *X* of some degree *a* grounds possession of ability *Y* to achieve value *e* in some continuous range between 0 and *r*; and $r > 0$ (e.g., some [any] ability to achieve ethical value).
3. The magnitude of range *r* varies as a function of the size of *a* (i.e., consistent with the pressure of reason).
4. For any human with the ability *Y* with range *r*, that human has the potential to achieve any particular value of *e*, between 0 and *r*.
5. There is a further value *m* calculated such that $m = e/r$ (e.g., moral value).
6. For any value of *e* and *r*, *m* will vary between 0 and 1 (i.e., between 0 percent and 100 percent).
7. Thus, for all humans, each with any degree *a* of nonnormative scalar property *X* grounding some (any) degree of ability *Y* to achieve value *e* in a range between 0 and any *r*, each human will also have the same (equal) ability *Z* to achieve *m* between 0 and 1 (e.g., equal ability to attain moral value).
8. Given 7, since those humans who possess nonnormative scalar property *X* do so naturally and inalienably, then those humans will also possess equal ability *Z* naturally and inalienably.
9. For any set of humans, the fact that those humans possess equal ability *Z* has implications of great weight or lexical priority for how they should be treated (e.g., it is fitting that those with equal ability to attain moral value should have the equal ability to attain ethical value).
10. Thus, given 8 and 9, *Z* is a basic moral property (as per the definition in section 1).
11. *Basic Equality*: All (or nearly all) humans are one another's equals in their possession of some basic moral property (i.e., *by virtue* of possessing some (any) degree of nonnormative property *X*).

67 In the loose sense of "natural" and "inalienable" defined in section 1.

6. Conclusion

This paper began by clarifying the distinction between various concept(ions) of basic equality and the agnostic proposition Basic Equality—that is, all (or nearly all) humans are one another's equals in their possession of some basic moral property. It then argued that the problem of basic equality is really an equally agnostic argument against Basic Equality: the pressure of reason argument. Ever since John Rawls's rather cursory reflection on the basis of basic equality, however, most theorists of basic equality have taken his range property strategy to offer the best possible hope of solving its problem. However, this paper has argued that while a range property grounded on an underlying nonnormative scalar property that we might share with other beings (like rational capacities) is, indeed, likely to be an incidental output of identifying a further nonnormative property that we (most likely) only share with (at least almost) all other human beings (like rational agency), the really hard task of avoiding the pressure of reason remains: Why do the scalar degrees of this further nonnormative property not then ground proportional scalar degrees of any basic moral property it grounds? While other theorists after Rawls have implicitly adopted a cognate chained bare property strategy to counter this riposte, I have argued that it still fails. Yet hope lies in the remainder. As Herbert Spiegelberg tangentially illustrated, it is possible that individuals with abilities of different scale can still have a kind of equality between them: one of equal relative potential. Much would be needed to flesh out and justify such an approach, but *prima facie*, the debate would be moving to new terrain, overcoming the pressure of reason problem to instead focus on the foundational moral implications, if any, of equal relative potential.[68]

University of Glasgow
nikolas.kirby@glasgow.ac.uk

References

Arendt, Hannah. *On Revolution.* New York: Penguin Books, 2006.

Arneson, Richard. "Basic Equality: Neither Acceptable nor Rejectable." In Steinhoff, *Do All Persons Have Equal Moral Worth?*, 30–52.

———. "What (If Anything) Renders All Human Persons Morally Equal?" In

68 Many thanks to Jo Wolff, Ian Carter, Giacomo Floris, Annabelle Lever, and Richard Arneson for their feedback on various iterations of this paper and its central ideas, as well the participants in the MANCEPT "Basic Equality" Workshop, September 2020.

Singer and His Critics, edited by Dale Jamieson, 103–28. Oxford: Blackwell, 1999.

Bedau, Hugo Adam. "Egalitarianism and the Idea of Equality." In *Equality*, edited by J. Roland Pennock and John W. Chapman, 3–27. New York: Atherton Press, 1967.

Benn, Stanley. "Egalitarianism and the Equal Consideration of Interests." In *Equality*, edited by J. Roland Pennock and John W. Chapman, 61–78. New York: Atherton Press, 1967.

Bennett, Christopher. "Intrusive Intervention and Opacity Respect." In *Treatment for Crime: Philosophical Essays on Neurointerventions in Criminal Justice*, edited by David Birks and Thomas Douglas, 255–73. Oxford: Oxford University Press, 2018.

Berker, Selim. "The Explanatory Ambitions of Moral Principles." *Noûs* 53, no. 4 (December 2019): 904–36.

———. "The Unity of Grounding." *Mind* 127, no. 507 (July 2018): 729–77.

Carter, Ian. "Respect and the Basis of Equality." *Ethics* 121, no. 3 (April 2011): 538–71.

Christiano, Thomas. *The Constitution of Equality: Democratic Authority and Its Limits*. Oxford: Oxford University Press, 2008.

———. "Rationality, Equal Status, and Egalitarianism." In Steinhoff, *Do All Persons Have Equal Moral Worth?*, 53–75.

Dworkin, Ronald. *Sovereign Virtue: The Theory and Practice of Equality*. Cambridge, MA: Harvard University Press, 2000.

———. *Taking Rights Seriously*. New impression. London: Duckworth, 1978.

Fine, Kit. "Guide to Ground." In *Metaphysical Grounding: Understanding the Structure of Reality*, edited by Fabrice Correia and Benjamin Schnieder, 37–80. Cambridge: Cambridge University Press, 2012.

Floris, Giacomo. "Two Concerns about the Rejection of Social Cruelty as the Basis of Moral Equality." *European Journal of Political Theory* 19, no. 3 (July 2020): 408–16.

Frankena, William K. "The Concept of Social Justice." In *Social Justice*, edited by Kenneth E. Boulding and Richard B. Brandt, 1–32. Englewood Cliffs: Prentice-Hall, 1962.

Frankfurt, Harry G. "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy* 66, no. 23 (December 1969): 829–39.

———. *On Inequality*. Princeton, NJ: Princeton University Press, 2015.

Gosepath, Stefan. "On the (Re)Construction and Basic Concepts of the Morality of Equal Respect." In Steinhoff, *Do All Persons Have Equal Moral Worth?*, 124–41.

Husi, Stan. "Why We (Almost Certainly) Are Not Moral Equals." *Journal of*

Ethics 21, no. 4 (December 2017): 375–401.
Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. In *Practical Philosophy*, translated and edited by Mary J. Gregor, 37–108. Cambridge: Cambridge University Press, 1996.
Kirby, Nikolas. "Two Concepts of Basic Equality." *Res Publica* 24, no. 3 (August 2018): 297–318.
Kymlicka, Will. *Contemporary Political Philosophy: An Introduction.* 2nd ed. Oxford: Oxford University Press, 2002.
Lucas, J. R. "Against Equality." *Philosophy* 40, no. 154 (October 1965): 296–307.
Macdonald, Margaret. "Natural Rights." *Proceedings of the Aristotelian Society* 47 (1946–47): 225–50.
McLaughlin, Brian P. "Varieties of Supervenience." In *Supervenience: New Essays*, edited by Elias E. Savellos and Umit D. Yalcin, 16–59. Cambridge: Cambridge University Press, 1995.
McMahan, Jeff. "Challenges to Human Equality." *Journal of Ethics* 12, no. 1 (January 2008): 81–104.
Miller, David. *National Responsibility and Global Justice.* Oxford: Oxford University Press, 2007.
Morton, Justin. "Grounding the Normative: A Problem for Structured Non-Naturalism." *Philosophical Studies* 177, no. 1 (January 2020): 173–96.
Parr, Tom, and Adam Slavny. "Rescuing Basic Equality." *Pacific Philosophical Quarterly* 100, no. 3 (September 2019): 837–57.
Pojman, Louis P. "A Critique of Contemporary Egalitarianism: A Christian Perspective." *Faith and Philosophy* 8, no. 4 (October 1991): 481–504.
———. "On Equal Human Worth: A Critique of Contemporary Egalitarianism." In Pojman and Westmoreland, *Equality*, 282–99.
Pojman, Louis P., and Robert Westmoreland. *Equality: Selected Readings.* New York: Oxford University Press, 1997.
Rawls, John. *A Theory of Justice.* Rev. ed. Cambridge, MA: Belknap Press, 1999.
Rosen, Gideon. "Metaphysical Relations in Metaethics." In *The Routledge Handbook of Metaethics*, edited by Tristram Colin McPherson and David Plunkett, 151–69. New York: Routledge, 2018.
Sangiovanni, Andrea. *Humanity without Dignity: Moral Equality, Respect, and Human Rights.* Cambridge, MA: Harvard University Press, 2017.
Schaar, John H. "Some Ways of Thinking about Equality." *Journal of Politics* 26, no. 4 (November 1964): 867–95.
Sher, George. *Me, You, Us: Essays.* New York: Oxford University Press, 2017.
Spiegelberg, Herbert. "A Defense of Human Equality." *Philosophical Review* 53, no. 2 (March 1944): 101–24.
Steinhoff, Uwe. "Against Equal Respect and Concern, Equal Rights, and

Egalitarian Impartiality." In *Do All Persons Have Equal Moral Worth?*, 142–72. Oxford: Oxford University Press, 2015.

———, ed. *On "Basic Equality" and Equal Respect and Concern*. Oxford: Oxford University Press, 2014.

Thomas, D. A. Lloyd. "Equality within the Limits of Reason Alone." *Mind* 88, no. 352 (October 1979): 538–53.

Vallentyne, Peter. "Of Mice and Men: Equality and Animals." *Journal of Ethics* 9, nos. 3/4 (2005): 403–33.

Vlastos, Gregory. "Justice and Equality." In *Theories of Rights*, edited by Jeremy Waldron, 41–76. Oxford: Oxford University Press, 1984.

Waldron, Jeremy. *One Another's Equals: The Basis of Human Equality.* Cambridge, MA: Belknap Press, 2017.

Westen, Peter. *Speaking of Equality: An Analysis of the Rhetorical Force of "Equality" in Moral and Legal Discourse.* Princeton Legacy Library. Princeton, NJ: Princeton University Press, 2014.

Williams, Bernard. *In the Beginning Was the Deed: Realism and Moralism in Political Argument.* Princeton, NJ: Princeton University Press, 2008.

———. *Problems of the Self.* Cambridge: Cambridge University Press, 1973.

Wilson, John. *Equality.* Philosophy at Work. London: Hutchinson, 1966.

Wolf, Susan. "The Importance of Free Will." *Mind* 90, no. 359 (July 1981): 386–405.

Journal of Ethics and Social Philosophy
VOL. 26, NO. 3 · FEBRUARY 2024
https://doi.org/10.26556/jesp.v26i3.1770

WHAT TIME TRAVEL TEACHES US ABOUT MORAL RESPONSIBILITY

Taylor W. Cyr and Neal A. Tognazzini

PHILOSOPHERS these days tend to favor ecumenical theories. It would be an undesirable feature of a theory of moral responsibility, for example, if it committed its proponents to a consequentialist theory of normative ethics. Likewise, it would be undesirable if a response to the problem of induction committed its proponents to theism. And so on.

The implicit acceptance of this methodological constraint opens up fruitful avenues of research for those inclined to see how a theory in one area of philosophy might have consequences for theorizing in another area. In this paper, we would like to explore one of these avenues. Specifically, taking our cue from a recent paper by Yishai Cohen, we would like to see what the metaphysics of time travel might be able to teach us about moral responsibility.[1] In his paper, Cohen argues that if time travel is metaphysically possible, then one of the most influential theories of moral responsibility—that of John Martin Fischer and Mark Ravizza—is false.[2] If Cohen were right, that would be an especially surprising connection between literatures that have, for the most part, developed independently of each other.[3]

In what follows, we will argue that Cohen is right to think that we can learn something important about moral responsibility from the metaphysics of time travel but that the true lesson is not quite the one he has in mind. In particular, we will show that although Cohen's argument is unsound, it can nevertheless serve as a lens to bring reasons-responsive theories of moral responsibility into sharper focus, which in turn will help us to better understand *actual-sequence* theories of moral responsibility more generally.

1 Cohen, "Reasons-Responsiveness and Time Travel."

2 Fischer and Ravizza, *Responsibility and Control.*

3 Spencer, "What Time Travelers Cannot Not Do," and McCormick, "A Dilemma for Morally Responsible Time Travelers," are notable exceptions to this generalization.

1

What connects the metaphysics of time travel with theories of moral responsibility are *counterfactuals*. So, let us begin by tracing both topics to their meeting point.

Moral responsibility is often thought to require free will, and free will is often thought to require the ability to do otherwise. Further, the ability to do otherwise is often thought to imply the truth of certain counterfactual claims. Take, for example, the infamous and discredited conditional analysis, according to which someone is able to do otherwise just in case, were they to desire to do otherwise, they would. Here, free will is analyzed in terms of a particular counterfactual.

But even theorists who endorse Harry Frankfurt's attack on the Principle of Alternative Possibilities—that is, even theorists who deny that moral responsibility requires the ability to do otherwise—still often talk about moral responsibility in terms of counterfactuals.[4] Take, for example, the most detailed and influential theory of moral responsibility on the market: that of John Martin Fischer and Mark Ravizza.[5] Fischer and Ravizza deny that moral responsibility requires the ability to do otherwise; instead, they offer an *actual-sequence* account of moral responsibility, according to which when an agent is morally responsible, this is wholly in virtue of facts about the way an action is *actually* produced, and not at all in virtue of facts about how things *might* have unfolded or *would* have unfolded in some non-actual possible world. But which actual-sequence facts matter for moral responsibility?

Fischer and Ravizza focus their attention on the so-called *control condition* for moral responsibility (as opposed to, say, the *epistemic* condition, which is also important but not as frequently discussed), and their contention is that an agent has control over what they do just in case their action issues from their *own, moderately reasons-responsive mechanism*. We will get into some of the details of their account below, but for now, it suffices to note that despite their being champions of an actual-sequence account of moral responsibility, Fischer and Ravizza still rely heavily on counterfactuals in spelling out the notion of reasons-responsiveness. Instead of focusing on what the *agent* would do under certain counterfactual circumstances, however, they focus on the reasons-sensitivity of the agent's *decision-making mechanism*, where that mechanism is sensitive to reasons just in case certain counterfactuals hold. This is a subtle argumentative strategy, and it is, of course, not without its share of

4 Frankfurt, "Alternate Possibilities and Moral Responsibility."

5 Fischer and Ravizza, *Responsibility and Control*.

critics; but again, we will save some of the details for later. For now, the point is that theorizing about moral responsibility seems to lead inevitably to a careful consideration of certain counterfactuals.[6]

The same can be said for the metaphysics of time travel. Here the connection is even easier to see since philosophical discussions about time travel have tended to center around the Grandfather Paradox and other similar worries about the possibility of backward time travel. Briefly, the worry is that if backward time travel is possible, then contradictions could be true. The rough idea is as follows: if backward time travel is possible, then I could travel back in time to visit my grandfather when he was a child, and in that moment, it would be true *both* that I could kill him—what would stop me?—and also that I *could not* kill him—since if he had died in that moment, my mother would never have been born, and then I would never have existed, so I would not be there trying to decide what to do in the first place. The fact that I am there in his childhood means he did not die in that moment, so it looks like no matter how hard I try to kill him, I will inevitably fail, despite the fact that I have everything I would need in order to pull it off.

This is a rough-and-ready presentation of the paradox, so let us not put too much weight on it.[7] The relevant point is that a proper articulation and evaluation of the Grandfather Paradox will require a deep dive into counterfactual reasoning. For example, the scenario sketched above seems problematic in part because it seems to be describing a situation in which the following counterfactual is true: if I were to kill my grandfather, then I would not have existed. That by itself seems to cause trouble for the supposition that I *can* kill my grandfather while I am time traveling, but we can cause even more trouble for that supposition by endorsing the following principle, inspired by Kadri Vihvelin: *S* is able to do *A* only if, had *S* tried to do *A*, *S* would or at least might have succeeded.[8] The funny thing about me and my grandfather is that, *no matter how hard I were to try*, I would fail to kill him. And if the principle just mentioned is correct, then it follows that I *cannot* kill him.

One of the perplexing things about backward time travel—at least, cases of it that involve the time traveler visiting their past self or their direct ancestors—is that it makes counterfactuals go all screwy. All of a sudden, my own existence appears to hinge on (i.e., counterfactually depend on) the most mundane of events. Parricide is not mundane, of course, but that is just a particularly vivid

6 There is an important exception to this claim that we discuss in section VI below.

7 See Wasserman, *Paradoxes of Time Travel*, chs. 3 and 4, for a comprehensive discussion of this and related paradoxes.

8 Vihvelin, "What Time Travelers Cannot Do," 318.

example. In the *Back to the Future* film franchise, the same basic paradox is explored without parricide and instead with the simple and accidental event of keeping one's own parents from ever falling in love. But whatever the details of the story, in cases of backward time travel, our usual method for evaluating counterfactual statements seems to lead us into trouble since facts about the future (that is, about the time traveler's personal past, before they got into the time machine) seem like they must be held fixed—and, to put it simply, we just are not used to doing that. It is the *past* that is fixed, while the future is *open*. But in cases of time travel, as David Lewis puts it, facts about the future "masquerade" as facts about the past.[9]

So far, we have explained how our two topics—moral responsibility and time travel—both require careful thinking about counterfactuals, but this falls short of the task we set ourselves in this section, which is to show how the topic of counterfactuals *connects* theorizing about moral responsibility with the metaphysics of time travel. Now that we have the backstory, we can make relatively quick work of that task.

Here is the bottom line: the most influential theory of moral responsibility understands the crucial notion of *control* in terms of the holding of certain counterfactuals that provide the details about whether (and to what extent) an agent's action-producing mechanism is sensitive to reasons, but in cases of backward time travel, counterfactuals that we ordinarily take to be boringly true turn out to be bewilderingly false (or else we have no idea what to say about them). What that means is that there will be time travel stories that will seem, at least at first glance, to provide counterexamples to this theory of moral responsibility. As we have seen, in cases of time travel, we can get counterfactuals about the behavior of agents to come out false, *seemingly without interfering with the intrinsic capacities of the agents in question*, and instead just by placing them in the right external circumstances. So, if your preferred theory of moral responsibility is both (a) committed to the truth of certain counterfactuals and (b) ostensibly concerned solely with an agent's intrinsic psychological capacities, then you probably cannot have both of those things at the same time.

In the next section, we will look closely at a detailed version of this worry, raised recently by Yishai Cohen against Fischer and Ravizza's theory of moral responsibility. Our contention will be that although Cohen's argument is unsound, taking it seriously will teach us something important about theories of moral responsibility more generally, especially ones that claim to focus exclusively on the *actual sequence*.

9 Lewis, "The Paradoxes of Time Travel," 151.

II

In a recent paper, Yishai Cohen claims if we add one seemingly harmless thesis to the theory of moral responsibility championed by Fischer and Ravizza, then that theory is inconsistent with the metaphysical possibility of time travel. This would be a very odd result, to say the least, but it would also be an unattractive result, especially to Fischer and Ravizza, who are explicitly concerned with constructing a theory of responsibility that does not hinge on "the arcane ruminations" of theoretical physicists (or, presumably those of metaphysicians, either).[10] Moreover, there is fairly wide consensus among contemporary metaphysicians that the usual objections to the metaphysical possibility of time travel fail, so it would be a mark against Fischer and Ravizza's theory if it required them to take a dissenting view.[11] Fortunately, Cohen's attempt to saddle Fischer and Ravizza with this result is unsuccessful. But before we explain why, let us take a closer look at Cohen's argument.

To see how Cohen's argument works, we need to explain the Fischer and Ravizza account of moral responsibility in a bit more detail. We have already said that Fischer and Ravizza offer an account of the control condition on moral responsibility, and that they lay out a set of necessary and sufficient conditions for an agent's exercising that sort of control. They call it *guidance control*, and their account runs as follows:

> An agent exercises *guidance control* over an action just in case the action issues from the agent's own moderately reasons-responsive mechanism, where a mechanism is moderately reason-responsive just in case it is regularly receptive to reasons and at least weakly reactive to reasons.[12]

The notions of *regular receptivity* and *weak reactivity* here are spelled out in terms of how the mechanism would respond in various counterfactual circumstances:

10 Fischer, *My Way*, 5.

11 As Cohen notes (in "Reasons-Responsiveness and Time Travel," 6n19), Dowe defends the metaphysical possibility of time travel ("The Case for Time Travel"), and Artzenius and Maudlin discuss its nomological possibility ("Time Travel and Modern Physics"). For the classic defense of the metaphysical possibility of time travel, see Lewis, "The Paradoxes of Time Travel." For a more recent (and the first book-length) defense of the metaphysical possibility of time travel, see Wasserman, *Paradoxes of Time Travel*.

12 This is our paraphrase of the account elaborated and defended in Fischer and Ravizza, *Responsibility and Control*. We are setting aside the ownership component of guidance control since this does not play a role in Cohen's argument, but see Fischer and Ravizza, *Responsibility and Control*, ch. 8, for their account of ownership.

> A mechanism is *regularly receptive to reasons* just in case there are possible scenarios in which (1) there is sufficient reason to do otherwise, the same kind of mechanism is operative, and the agent recognizes that reason, and (2) the possible scenarios described in 1 constitute an understandable pattern of reasons-recognition.
>
> A mechanism is *weakly reactive to reasons* just in case it is regularly receptive to reasons and, in at least one of the possible scenarios described in the account of regular receptivity, the agent chooses and does otherwise for the reason in question.[13]

These formulations are adequate, but they are also a bit abstract. Here is the basic idea: when a morally responsible agent acts, the process leading up to their action (the "mechanism") is capable of "seeing" the relevant reasons and is also capable of reacting appropriately to those reasons. To figure out whether a mechanism has the relevant capabilities, we look to facts about nearby worlds. So long as there is an intelligible range of possible circumstances in which this particular decision-making process *does* "see" the reasons there are, then we can say that the *actual* decision-making process is *capable* of "seeing" those reasons. Likewise, so long as there is at least one possible circumstance in which, having "seen" the reasons, the relevant decision-making process kicks into gear and issues in a choice on the basis of those reasons, then we can say that the *actual* decision-making process is *capable* of reacting to those reasons. (The rationale for why receptivity requires an "understandable pattern" whereas reactivity only requires "at least one" relevant possible scenario need not detain us here.)

One of Fischer and Ravizza's key innovations is to distinguish between the *agent* and the *mechanism* by which the agent acts.[14] They do this for two related reasons: (1) they are persuaded by so-called Frankfurt-style counterexamples that an agent can be morally responsible for what they have done even if the agent was not able to have done otherwise, and (2) they want to defend a positive theory of moral responsibility that focuses on the capacity to respond to reasons. Since the notion of capacity is a paradigm modal notion, Fischer and Ravizza need to find a way to get modality into their account without giving up on the insight of Frankfurt-style counterexamples. They do this by distinguishing between agents and mechanisms: the agent may not be able to do otherwise, but that does not mean the mechanism on which the agent acts is not *capable* of responding to the relevant reasons.

13 Again, we are paraphrasing. For the full details, see Fischer and Ravizza, *Responsibility and Control,* 69–76.

14 See Fischer and Ravizza, *Responsibility and Control,* 38–41.

But it is this very desire to accommodate a modal notion like *capacity* that, according to Yishai Cohen, puts the Fischer and Ravizza account of moral responsibility on a collision course with the metaphysical possibility of time travel. This is because, as we have seen, cases of backward time travel make trouble for our ordinary ways of thinking about counterfactuals. The essence of Cohen's objection is this: we can easily construct a backward time travel story according to which the time traveler seems for all the world to be morally responsible for what they have done, but, due to the metaphysical peculiarities involved in attempting to kill one's younger self, there does not exist the range of worlds that Fischer and Ravizza say is needed for the agent to be acting on a moderately reasons-responsive mechanism. Here is a modified version of the story that Cohen tells:

> Zoe lives in a peculiar world. First, time travel is nomologically possible. Second, individuals can commit murder merely by *willing that someone die*. However, there is one line of defense available to the would-be victims: they can continue to live simply by willing to nullify the attempted murder. Now, suppose that Zoe travels twenty years into the past to visit a younger version of herself, and suppose that she wills that her younger self die. However, her attempted murder does not succeed because her younger self wills to nullify the attempt.[15]

Now, Cohen claims that if we think carefully about the relationship that Zoe has to her younger self, we will see that the mechanism that the younger Zoe acts on cannot be moderately reasons-responsive. It is crucial that these are two person-stages of the very same individual because that means that the very existence of Zoe-the-time-traveler depends counterfactually on her failure to kill her younger self. With that in mind, we can see that once Zoe has become a time traveler, there are no worlds in which younger Zoe dies, and hence no worlds in which she refrains from willing to nullify her older self's attempted murder.[16] But if there are no worlds in which she refrains, then *a fortiori* there is not an "understandable pattern" of worlds in which she sees the reasons to refrain and then acts on them. But it is precisely this pattern of worlds that Fischer and Ravizza say is required for younger Zoe to be morally responsible for her behavior.

The argument is not yet complete, however. All that follows so far is that if Fischer and Ravizza are right about moral responsibility, then younger Zoe is not morally responsible for willing to nullify her older self's attempted murder.

15 Cohen, "Reasons-Responsiveness and Time Travel," 3.

16 This is a bit too quick, actually, since there may be worlds in which young Zoe is killed by her future self but is then somehow resurrected. (Thanks to Ryan Wasserman for discussion here.) We set these sorts of worries aside, however, since our aim is to draw lessons for theorizing about moral responsibility.

For this story to constitute a *worry* for Fischer and Ravizza, we need some independent reason to think that their view gives us the *wrong* verdict about younger Zoe's moral responsibility. To secure this result, Cohen appeals to the following principle:

> *Intrinsic Mechanism*: Whether a mechanism is moderately reasons-responsive depends only on the *intrinsic* properties of the agent in question.[17]

Cohen admits that Fischer and Ravizza do not explicitly endorse this principle, but he argues that it would be better, *ceteris paribus*, for them to accept it. And it certainly does have the ring of truth: after all, facts about the capacities of my decision-making processes do not seem to depend on anything happening across town. To know whether my capacities are reasons-responsive, it seems like you would only need to look at those capacities themselves.[18]

If we accept Intrinsic Mechanism, and we agree that the story of Zoe is metaphysically possible, then we can create a problem for Fischer and Ravizza. Recall that younger Zoe does not act from a moderately reasons-responsive mechanism since there are no worlds in which she refrains from acting in self-defense, and hence no worlds that can serve as witness to the claim that her decision-making process is responsive to reasons. But now just tweak Zoe's story a bit so that young Zoe does not face an older version of herself but instead faces a time traveler with no interesting counterfactual dependency on her—Cohen calls her "Amy." Notice that this tweak of the story does not alter any of *young Zoe's* intrinsic properties: all we have done is remove older Zoe from the story and replace her with a time traveler named Amy. But the second we break the counterfactual dependency between murderer and victim, we also get all the relevant possible worlds back in which young Zoe refrains from willing to nullify the attempted murder, which means that young Zoe miraculously becomes responsive to reasons again, despite our not having changed any of her intrinsic properties.

The upshot? If we accept Intrinsic Mechanism, then we have to say that young Zoe's mechanism is reasons-responsive in both stories or in neither, but the Fischer and Ravizza account is at odds with that verdict. According

17 This is our paraphrase of Cohen's principle: "A moderately reasons-responsive mechanism *M* that issues in *S*'s ϕ-ing is *wholly* constituted by *S*'s intrinsic properties (either all of *S*'s intrinsic properties or, more likely, some subset thereof)" ("Reasons-Responsiveness and Time Travel," 2).

18 While we can grant this claim for the sake of argument, Cohen's argument that Fischer and Ravizza should accept it is problematic. In particular, Cohen gives an example of one clearly irrelevant extrinsic property (being one mile away from a post office) and then claims that this suggests that only intrinsic properties are relevant to reasons-responsiveness. But this is a bit too quick; it would not follow from the irrelevance of one extrinsic property that all extrinsic properties are irrelevant.

to Fischer and Ravizza's account, whereas young Zoe is not moderately reasons-responsive in the version of the story where she confronts her older self, young Zoe *is* moderately reasons-responsive in the version of the story where she confronts Amy (or, at least, there is no reason in the Amy story to think that young Zoe *is not* moderately reasons-responsive). Something has to go, and since *Intrinsic Mechanism* is the most plausible of the bunch, the worry here can be adequately framed as a conflict between the metaphysical possibility of time travel and the Fischer and Ravizza account of moral responsibility.

III

We have three worries about Cohen's objection. The first worry shows that his objection, even if successful, is more limited in scope than it at first seems. The second two worries show that even the limited objection fails.

First: although Cohen describes his conclusion as the claim that Fischer and Ravizza's account of moral responsibility is incompatible with the metaphysical possibility of time travel, nothing quite so grand follows from the considerations he adduces, even if his arguments are sound. Rather, all that would follow is that *the time travel stories involving Zoe and Amy* are incompatible with the Fischer and Ravizza account of moral responsibility. Of course, we could generalize a further conclusion by abstracting away from the particular imaginary individuals in those stories, but still, at best, that would give us the claim that the metaphysical possibility of *single-timeline backward time travel involving agents* is incompatible with the Fischer and Ravizza account of moral responsibility. This is not an insignificant conclusion since these are precisely the sorts of time travel stories that tend to capture the imaginations of sci-fi lovers. Still, single-timeline models of time travel are not the only feasible models, backward is not the only direction one might wish to travel, and, in the actual world at least, non-agential travel through time would probably be the first breakthrough to make headlines. So, Cohen's conclusion is more limited than advertised.

Even thus qualified, though, there are two major problems with Cohen's argument. The first is that Cohen does not respect the distinction that Fischer and Ravizza draw between *agents* and their *mechanisms*. The second is that Cohen fails to appreciate the significance of Fischer and Ravizza's claim that reactivity is "all of a piece," so that if a mechanism can react to *any* reason to do otherwise, then it can react to *all* such reasons.[19] We will take these two problems in order.

19 As an anonymous reviewer points out, if we consider a view like Fischer and Ravizza's but that lacks these two features (the distinction between agents and mechanisms and the claim that reactivity is all of a piece), such a view *would* fall prey to certain time-travel

First, consider one more time why younger Zoe seems not to be acting from a moderately reasons-responsive mechanism when she faces off against her older, time-traveling self. Although what Zoe *actually* does is will to nullify the attempted murder, in order for that to be an action for which she is morally responsible, there must be a suitable range of worlds in which Zoe recognizes reasons to refrain from nullifying the attempted murder, and there must be at least one world in which, having recognized those reasons, Zoe *does* refrain from nullifying the attempted murder. But since the would-be murderer is her older self, we know that there are *no* worlds in which she refrains from nullifying the attempted murder. Hence, younger Zoe's nullifying actions cannot have issued from a reasons-responsive mechanism.

But if you look closely at the justification just offered, you will see that we have moved back and forth between talking about Zoe herself, on the one hand, and talking about Zoe's action-producing mechanism, on the other. And in fact, the justification gains whatever superficial plausibility it has precisely from this equivocation. On Fischer and Ravizza's official account, everything is done in terms of *mechanisms* rather than *agents*. So, in order to get the same result—that younger Zoe is not acting from a reasons-responsive mechanism when she nullifies her older self's attempted murder—we have to show, not that there are no worlds in which *Zoe* refrains from the act of nullifying, but rather that there are no worlds in which *her mechanism issues in an act of refraining*. It is the mechanism, after all, which has (or does not have) the property of being responsive to reasons, and the agent acquires that status only derivatively.

Paying close attention to the difference between agents and mechanisms helps us to see how Fischer and Ravizza can escape Cohen's criticism. The feature of the time travel example that is so peculiar is that the person attempting murder and the person who is the victim of an attempted murder *are the same person*—this is why it does not make sense to imagine a world in which young Zoe fails to stop her own murder (i.e., a world in which she dies at the hands of her future self). But the Fischer and Ravizza account of moral responsibility does not apply at the level of persons—at least, not in the first instance. Instead, it applies at the level of *mechanisms*. And there is nothing contradictory about saying that the relevant mechanism might have issued in some other willing since we need not hold everything fixed about the agent whose mechanism it is in order to figure out what capacities the mechanism itself has.

---

scenarios (though Cohen's argument would still need to be qualified in the way we indicated above). But, as far as we know, no one holds such a view, and we are interested in defending Fischer and Ravizza's account. Perhaps, though, Cohen's challenge to Fischer and Ravizza serves to highlight the importance of these two features of the account.

Perhaps another way to put the point is to say that whereas there are no possible worlds in which young Zoe fails to stop *her own murder at the hands of her future self*, there certainly are possible worlds in which the type of mechanism on which young Zoe acts issues in the decision to let herself be killed. It is just that in those worlds, some of the external circumstances would have to be different. In those circumstances—the ones that we look to in order to figure out whether young Zoe's actually operative mechanism is responsive to reasons—perhaps the person attempting to murder her is an enemy combatant in a war, and she willingly sacrifices herself for the good of her community. There is, after all, no contradiction in the supposition that the mechanism on which young Zoe acts when she thwarts her older self's plan might nevertheless be the same kind of mechanism that, in a different circumstance, issues in a decision to sacrifice herself. (It is not as though young Zoe is *invincible*, after all.)

So, to sum up our first response to Cohen's objection: although there are no worlds in which young Zoe allows her older self to murder her, there are (it seems) plenty of worlds in which the relevant action-producing mechanism issues in a self-sacrificial decision due to the presence of different incentives. And it is this latter fact that tells us something about Zoe's moral responsibility, according to Fischer and Ravizza.

IV

The second reason why Cohen's objection fails has to do with a rather peculiar claim that Fischer and Ravizza make about the notion of *weak reasons-reactivity*. If you look back at the account of guidance control that we spelled out above, you will notice that guidance control involves both *receptivity* and *reactivity*, but whereas Fischer and Ravizza classify the relevant sort of receptivity as *regular*, they classify the relevant sort of reactivity as *weak*. And indeed, when they spell out what those terms mean, we can see that they correspond to different spheres of possible worlds. A mechanism is *regularly receptive* to reasons just in case there is an intelligible pattern of counterfactual circumstances in which the mechanism would "see" the reasons at play, but a mechanism is *weakly reactive* to reasons just in case there is *at least one* counterfactual circumstance in which the mechanism would respond to those reasons, upon seeing them. Why the asymmetry?

Fischer and Ravizza opt for *weak* reasons-reactivity because, as they put it, reactivity is "all of a piece."[20] Here is what they mean: "If an agent's mechanism reacts to *some* incentive to (say) do otherwise than he actually does, this

20 Fischer and Ravizza, *Responsibility and Control*, 73.

shows that the mechanism *can* react to *any* incentive to do otherwise."[21] This is meant to mark a crucial difference between receptivity and reactivity. When it comes to receptivity, Fischer and Ravizza are worried about the possibility of a responsibility-undermining sort of blind spot in moral reasoning. They think it is possible, for example, that you might be able to recognize the fact that your action would break a promise as a reason not to do it, and yet you might not be able to recognize the fact that your action would cause me pain as a reason not to do it. That is, they are worried about mechanisms that are pathological in such a way that although certain moral reasons are on their radar, other moral reasons that seem like they should be equally visible just are not on their radar. Such a person, Fischer and Ravizza maintain, ought to be excused due to this bizarre malfunction in receptivity.

But when it comes to "the capacity to *translate* reasons into choices (and then subsequent behavior)"—that is, when it comes to the capacity that Fischer and Ravizza call "reactivity"—their claim is that such a bizarre sort of "blind spot" is impossible.[22] In fact, it would not even be right to call it a "blind spot" in this instance since we are talking about *reactivity* rather than *receptivity*. So, the "all of a piece" claim is that, so long as your mechanism would react *at all*—so long as it is "online," so to speak—then it does not matter what precise reason we put into the mechanism. If there is a scenario in which it reacts to *one* reason, then it has the capacity to react to them *all*. And that is why we only need to look at *one* possible world to determine whether a mechanism is appropriately *reactive* to reasons, even though we need to look at a suitably wide range of worlds to determine whether a mechanism is appropriately *receptive* to reasons.[23]

We have tried to keep the details to a minimum here, but they are important for seeing where Cohen's criticism goes wrong. Recall again why we are supposed to think that young Zoe fails to meet the criteria for exercising guidance control: given the peculiarities involved in backward time travel, there is no world in which young Zoe fails to stop her own murder, and this shows us that the mechanism on which she acts is insensitive to reasons. This is a point about

21 Fischer and Ravizza, *Responsibility and Control*, 73.

22 Fischer and Ravizza, *Responsibility and Control*, 69.

23 As an anonymous reviewer points out, Fischer and Ravizza's claim that reactivity is all of a piece seems to count as morally responsible some extremely weak-willed agents (e.g., a severe drug addict) who are intuitively not morally responsible, since there may well be one, possibly outlandish, scenario in which even a weak-willed agent's mechanism reacts successfully. Fischer and Ravizza explicitly acknowledge this implication of their view in their discussion of Brown and the drug "Plezu" (*Responsibility and Control*, 73–74). In later work, responding to an objection from Mele, "Reactive Attitudes, Reactivity, and Omissions," Fischer tentatively suggests that we might say that such an agent is morally responsible but not blameworthy (*Deep Control*, 187–92). For further discussion, see Cyr, "Semicompatibilism," 315.

reactivity: there is no possible scenario in which the relevant mechanism *reacts* to the reasons there may be for refraining from nullifying the murderous action since "reaction" is a matter of translating reasons into choices and behavior, and of course, it is not possible for young Zoe to be killed by her older self. So, it looks as though Zoe's mechanism does not have the sort of reactivity that Fischer and Ravizza think is needed for guidance control.

But again, this reasoning relies on a sort of equivocation. This time the equivocation is not between *agent* and *mechanism* but instead between two sorts of *reason* to which the mechanism might react. If we focus just on the *exact reason* that young Zoe acts on—namely, one that makes essential reference to the peculiar situation she finds herself in, where her older self is trying to kill her—then Cohen is right to say that there is no possible scenario in which the mechanism reacts differently to *that reason*. However, this is not enough to show that the mechanism fails to satisfy the reactivity criterion on guidance control because remember: according to Fischer and Ravizza, reactivity is *all of a piece*. So long as there is *at least one* possible scenario in which young Zoe's mechanism successfully reacts to a reason *of the same sort* as the one we are wondering about, then that is sufficient for us to conclude that the mechanism is *capable* of reacting to the reason we are wondering about.

In order to figure out whether young Zoe's mechanism is appropriately reactive to reasons, then, we do not need to find a scenario in which she fails to stop *her older self* from killing her. Instead, we just need to find a scenario in which she fails to stop *someone* from killing her. We just need to know whether the reason in question is the *sort* of reason her mechanism is able to translate into action, not whether there is a genuine possibility that *this particular* reason gets translated into action.

Cohen considers an objection along these lines, that perhaps all we need to know about the mechanism is that it is capable of reacting to a threat from some "different but qualitatively similar" person. Cohen's response is to say that "even if there is a nomologically identical world in which Young Zoe refrains . . . from nullifying the act of someone who is qualitatively similar to Old Zoe, this has no bearing upon whether [Young Zoe's mechanism] is moderately reasons-responsive."[24] But this response fails to appreciate the claim that reactivity is "all of a piece." This claim is precisely what allows us to move from "possibly, young Zoe's mechanism reacts to a reason of the same sort" to "actually, young Zoe's mechanism is capable of reacting to the actual reason."[25]

24 Cohen, "Reasons-Responsiveness and Time Travel," 5.

25 For discussion of Fischer and Ravizza's claim that reactivity is "all of a piece," and for a potential worry given that receptivity is *not* "all of a piece," see Todd and Tognazzini, "A Problem for Guidance Control."

V

The last two sections have gotten us pretty far into the weeds, and that is because Cohen's objection focuses specifically on the Fischer and Ravizza theory of moral responsibility, which has been worked out in great detail. To construct an adequate response on behalf of Fischer and Ravizza, then, we have had to look at those details. But now we want to zoom out a bit. First, we will give a high-altitude summary of why Cohen's objection fails. But then we will try to articulate what we think is insightful about Cohen's worry and what implications that insight has for theorizing about moral responsibility more generally. In the end, we will see that this will help us to bring even Fischer and Ravizza's account into sharper focus.

So, first, here is the high-altitude summary of our reply to Cohen. Cohen's basic worry is as follows: if moral responsibility is a matter of reasons-responsiveness, then merely changing an agent's external circumstances should not make a difference to whether they are morally responsible. But cleverly constructed time travel examples can screw up counterfactuals about an agent without changing anything intrinsic to the agent herself, so if we understand reasons-responsiveness in terms of counterfactuals, then we will be able to eliminate moral responsibility merely by changing an agent's external circumstances. Hence there is a deep tension at the heart of the Fischer and Ravizza account. On the one hand, they want reasons-responsiveness to be a matter of an agent's *intrinsic* properties, but on the other hand, they want to understand reasons-responsiveness in terms of *counterfactuals*. And what time travel stories show us (among other things) is that counterfactuals about an agent can vary independently of the agent's intrinsic properties, so it looks like Fischer and Ravizza cannot have both of the things they want.

Our basic reply is to say that Cohen has been looking at the wrong counterfactuals. Time travel examples involving retro-suicide attempts do mess up counterfactuals about the *agent*, but it is not clear that they mess up counterfactuals about the *mechanism*. (This was our first substantive reply.) Moreover, even if time travel examples show that there is no way the mechanism will react to the *actual reason*, that does not show that the mechanism *cannot* react to the actual reason since reactivity is all of a piece. All Fischer and Ravizza need is the claim that there is some reason of the same sort that the mechanism possibly reacts to. (This was our second substantive reply.)

But even if Cohen's objection fails, there is likely to be a lingering worry here, which might be expressed rhetorically as a question: Why exactly is an *actual-sequence* account of moral responsibility trafficking in *counterfactuals* in the first place? Facts about what *could* have or *would* have happened seem

like the basic ingredients of a theory of moral responsibility that emphasizes *alternative possibilities*. True, Fischer and Ravizza make the move from talking about what an *agent* can do to talking about what a *mechanism* can do (or is capable of doing), but this move might seem a bit like cheating since it seems to smuggle alternative possibilities in through the back door.[26] Cohen's objection is made possible by the fact that Fischer and Ravizza emphasize the importance of counterfactuals, and yet we can use time travel to generate some surprising counterfactual results. Although the objection fails, it provides the occasion to rethink the framing of Fischer and Ravizza's view since—in our view—counterfactuals ought not to have a prominent place in an actual-sequence theory of moral responsibility in the first place.

VI

We are not the first to note the awkwardness of being committed to an *actual-sequence* account of moral responsibility but yet giving pride of place to *counterfactuals* in the details of that theory. This criticism has also been raised forcefully by Christopher Franklin in his descriptively titled paper, "Everyone Thinks That an Ability to Do Otherwise Is Necessary for Free Will and Moral Responsibility."[27] According to Franklin, despite their claim to be providing an actual-sequence account of moral responsibility, Fischer and Ravizza's account requires alternative possibilities after all. As we have seen, Fischer and Ravizza's account of guidance control includes the following reactivity component:

> A mechanism is *weakly reactive to reasons* just in case it is regularly receptive to reasons and, in at least one of the possible scenarios described in the account of regular receptivity, the agent chooses and does otherwise for the reason in question.

In order for an agent to be morally responsible, then, the agent's operative mechanism must react to a reason to do otherwise in some possible scenario. But this is just to say that the mechanism can do (or is capable of doing) otherwise, which is tantamount to saying that the mechanism has alternative possibilities. In Franklin's words, Fischer and Ravizza are committed to the view

26 It has seemed that way to many commentators. See, for example, Watson, "Reasons and Responsibility," 382.

27 Franklin, "Everyone Thinks That an Ability to Do Otherwise Is Necessary for Free Will and Moral Responsibility."

that "a mechanism is appropriately reactive only if it has certain dispositions or abilities, namely the ability to act on different sufficient reasons."[28]

Again, in order to preserve the insight of Frankfurt-style counterexamples, Fischer and Ravizza aim to show that morally responsible agents need not be able to do otherwise or have alternative possibilities, even though the account does require that morally responsible agents act from a weakly reactive mechanism. As Franklin argues, however, what is true of agents' mechanisms holds for agents themselves too:

> Agents make choices, act, and are morally responsible in virtue of the activity of their mechanisms. . . . If the agent's mechanism is able to do otherwise, then the agent is, in virtue of taking responsibility for the mechanism, able to do otherwise. A central contention, therefore, of Fischer [and Ravizza]'s theory of moral responsibility is that agents are morally responsible only if they possess an ability to do otherwise.[29]

If Franklin is right, then why do Fischer and Ravizza deny that morally responsible agents must have the ability to do otherwise? Franklin says that it is because Fischer and Ravizza really intended (or at least should have intended) to say that "certain *species* of abilities are irrelevant"), specifically the sort of ability that agents in Frankfurt-style counterexamples *lack*.[30] But once we distinguish that sort of ability from the ability required by the reactivity component of Fischer and Ravizza's account, it is clear that the account does require *some* ability to do otherwise.

Now, we think that Franklin's criticism fails because he has conflated an ability to do otherwise with the mere presence of "alternative possibilities."[31] It is true that Fischer and Ravizza look to possible worlds in order to determine whether an agent's mechanism is suitably reasons-responsive, but it does not follow from the modal facts themselves that an agent who acts from a suitably reasons-responsive mechanism is thereby *able* to have done otherwise. To have an ability requires more than the possession of just any alternative possibility.

28 Franklin, "Everyone Thinks That an Ability to Do Otherwise Is Necessary for Free Will and Moral Responsibility," 2096.

29 Franklin, "Everyone Thinks That an Ability to Do Otherwise Is Necessary for Free Will and Moral Responsibility," 2097.

30 Franklin, "Everyone Thinks That an Ability to Do Otherwise Is Necessary for Free Will and Moral Responsibility," 2097, emphasis added. This is related to the distinction some authors draw between "general" and "specific" abilities. See, for example, Mele, "Agents' Abilities"; and Whittle, "Dispositional Abilities."

31 A detailed version of this response to Franklin can be found in Cyr, "Semicompatibilism." See also Kittle, "Does Everyone Think the Ability to Do Otherwise Is Necessary for Free Will and Moral Responsibility?"

For example, it may be that getting a hole in one is a genuinely possible alternative to my hitting the bunker, but (trust me) I do not have the ability to hit a hole in one. Still, one might think that the spirit behind Franklin's criticism survives this response. The lesson we are supposed to have learned from Frankfurt-style counterexamples, one might think, is that facts about other worlds are simply *irrelevant* to whether an agent is actually morally responsible for what they do. And so there appears to be a sense in which Fischer and Ravizza—those great champions of Frankfurt-style compatibilism—have misunderstood the central lesson of the examples.

But a lot depends here on what is meant by the term 'irrelevant.' As the literature on Frankfurt-style counterexamples and semicompatibilism developed late last century, the main question was whether an ability to do otherwise was *necessary* for moral responsibility. Actual-sequence theorists said no, whereas leeway theorists said yes. Over the last twenty years, however, philosophers have more carefully distinguished between "mere" necessary conditions of a claim, on the one hand, and factors *in virtue of which* a claim is true.[32] And that means that there are now *three* different views theorists might have on the question of how alternative possibilities relate to moral responsibility.

> *Necessary and Grounded In*: Someone's being morally responsible not only *entails* the presence of alternative possibilities but is *partly grounded in* the existence of those alternative possibilities.
>
> *Necessary but Not Grounded In*: Someone's being morally responsible *entails* the presence of alternative possibilities, but it is *not even partly in virtue of* those alternative possibilities that the person is morally responsible.
>
> *Neither Necessary nor Grounded In*: Someone's being morally responsible *neither entails nor is grounded in* facts about alternative possibilities.

Although Fischer and Ravizza were writing before the contemporary literature on grounding really took off, it is clear that their theory falls into the second of these three categories, and *this* is the sense in which it is an "actual-sequence" theory: although facts about other worlds follow from their account of reasons-responsiveness, it is not *in virtue of* those otherworldly facts that a mechanism is reasons-responsive.[33] Rather, those otherworldly facts are what

32 See, for example, Fine, "Ontological Dependence"; Correia, "Ontological Dependence"; and Clark and Liggins, "Recent Work on Grounding."

33 As a matter of historical interest, Frankfurt himself has clearly distinguished between necessary conditions for moral responsibility, on the one hand, and facts in virtue of which someone is morally responsible, on the other, and he agrees with Fischer here. Responding

they are *because* the actual-world mechanism is reasons-responsive. It is easy to conflate "direction of reasoning" with "direction of explanation," but they are crucially different. The otherworldly facts are reasons to believe that the actual mechanism is reasons-responsive, but they are not explanations of why it has that feature.[34]

After making their claim that "reactivity is all of a piece" (discussed above), for example, Fischer and Ravizza appeal to grounding:

> Our contention, then, is that a mechanism's reacting differently to a sufficient reason to do otherwise in some other possible world shows that the same kind of mechanism can react differently to the *actual* reason to do otherwise. This general capacity of the agent's actual-sequence mechanism—and *not* the agent's power to do otherwise—is what helps to ground moral responsibility.[35]

In more recent work, Fischer has again made this point quite explicit, conceding to Franklin that perhaps he could have been clearer in previous work. Fischer says:

> I completely agree with Franklin that I do indeed believe that various kinds of alternative possibilities are required for moral responsibility (although not for the "grounding" or explanation of moral

---

to a criticism of his argument against the Principle of Alternative Possibilities (PAP), Frankfurt says, "The critical issue concerning PAP, then, is not whether it is always possible that an agent who is morally responsible for performing a certain action might have acted differently. Rather, it is whether that possibility—even assuming that it is real—*counts for anything* in determining whether he is morally responsible for what he did. My claim is that it does not" ("Some Thoughts concerning PAP," 340, emphasis added). See also Leon and Tognazzini, "Why Frankfurt-Examples Don't Need to Succeed to Succeed," for an examination of how the difference between necessity and grounding ought to shape our understanding of Frankfurt-style counterexamples.

34 An anonymous reviewer points out that even if Fischer and Ravizza do not give the otherworldly facts a role in grounding an agent's responsibility, merely acknowledging that they follow from the presence of responsibility is enough to undermine Fischer and Ravizza's claim to be offering a semi-compatibilist account of moral responsibility. Semi-compatibilism is usually understood as the view that moral responsibility is compatible with determinism, regardless of whether determinism rules out the ability to do otherwise. But now if Fischer and Ravizza acknowledge that reasons-responsive mechanisms generate alternative possibilities, it looks like it *does* matter after all whether determinism rules out all alternative possibilities. But as we point out in the text just below, Fischer distinguishes the sort of alternative possibilities entailed by the presence of a reasons-responsive mechanism from the sort of ability to do otherwise that features in the official formulation of the semicompatibilist view.

35 Fischer and Ravizza, *Responsibility and Control*, 73.

> responsibility), and thus that my repeated contention that alternative possibilities are not required for moral responsibility might well have caused confusion.... But, as Franklin also notes, these were not the sorts of alternative possibilities I had in mind in contending that moral responsibility does not require alternative possibilities. I have absolutely no interest in showing that moral responsibility does not require general capacities or abilities to do otherwise, or various other kinds of abilities to do otherwise that abstract away from the particulars of the agent's history and/or present situation.... I have always been interested in the sort of alternative possibility that would be (or could plausibly be thought to be) ruled out by causal determinism. And, clearly, general abilities and indeed any sort of ability to do otherwise that abstracts away from features of the agent's past and/or current situation need not be inconsistent with causal determinism.[36]

So, even if Fischer's view implies that alternative possibilities are necessary for moral responsibility, and even if the view implies that *some* (general) ability to do otherwise is necessary for moral responsibility, Fischer maintains that these possibilities/abilities do not ground or explain moral responsibility.[37]

In this way, the theory of Fischer and Ravizza (as well as Fischer's more recent work) contrasts with two other sort of compatibilist views, the first of which takes the "neither/nor" option and the second of which takes the "both/and" option. Mesh theories like those inspired by Frankfurt and Watson offer accounts of moral responsibility according to which one need not even mention what happens in other worlds.[38] Frankfurt himself is explicit, in fact, that moral responsibility does not require reasons-responsiveness:

> I do not believe that the mechanism has to be reasons-responsive. The mechanism is constituted by desires and volitions and, in my view, what counts is just whether what the agent wills is what he really wants to will.... Someone who is wholeheartedly behind the desires that move him when he acts is morally responsible for what he does, in my judgment, whether or not he has any reasons for his deeds or for his desires.[39]

36 Fischer, "The Freedom Required for Moral Responsibility," 221.

37 Carolina Sartorio, *Causation and Free Will*, also opts for a version of compatibilism according to which facts about possible worlds are necessary but not part of what grounds an agent's moral responsibility. Sartorio goes one step further than Fischer and Ravizza, though, and claims that the otherworldly facts show us that *absences* are playing a causal role in the actual sequence.

38 Frankfurt, "Freedom of the Will and the Concept of a Person"; and Watson, "Free Agency."

39 Frankfurt, "Reply to John Martin Fischer," 28.

Now, perhaps a comprehensive account of "wholeheartedness" would need to appeal to otherworldly facts; we do not intend to take a stand on how best to spell out a mesh theory of the sort inspired by Frankfurt's work. The point is simply that, at least on the face of it, a mesh theory looks to be even more of an "actual-sequence" theory than a theory that emphasizes reasons-responsiveness. Whereas reasons-responsiveness theories entail facts about what agents are up to in other worlds, it is not clear that mesh theories do. They are similar, however, in rejecting the idea that an agent's moral responsibility is even partly *grounded* in those otherworldly facts.

However, there are compatibilist theories that take a "both/and" approach instead. Here we have in mind the view of the so-called new dispositionalists, who not only reject Frankfurt-style counterexamples but who also aim to give a positive view of free will in terms of dispositions, which are spelled out in counterfactual terms.[40] These are leeway compatibilists rather than source compatibilists, theorists who think that not only is an ability to do otherwise necessary for moral responsibility but also that one is morally responsible partly in virtue of such an ability. Even if Franklin is right that reasons-responsive theorists are aligned in an important way with leeway theorists—since they both develop theories that give pride of place to facts about other worlds—there is nevertheless an important difference between them since one seeks to explain moral responsibility in terms of those otherworldly facts, whereas the other seeks to explain moral responsibility only in terms of actual-sequence facts.

Fittingly, then, we have found another way in which the theory of Fischer and Ravizza is a *semicompatibilist* theory. The familiar sense of that term conveys the idea that determinism is compatible with moral responsibility, regardless of whether determinism rules out the ability to do otherwise. But now we have seen that Fischer and Ravizza also hold the view that moral responsibility is not even partly grounded in the presence of alternative possibilities, regardless of whether the facts that ground moral responsibility entail the existence of alternative possibilities. The first claim differentiates Fischer and Ravizza from leeway compatibilists like Vihvelin, whereas the second claim differentiates them from what we might say are "pure" actual-sequence compatibilists, such as Frankfurt.[41]

40 See, for example, Vihvelin, "Free Will Demystified" and *Causes, Laws, and Free Will*; and Fara, "Masked Abilities and Compatibilism." For a critique of these accounts, see Clarke, "Dispositions, Abilities to Act, and Free Will"; and Franklin, "Masks, Abilities, and Opportunities."

41 A wrinkle worth noting but not worth dwelling on: there is room for a theory of moral responsibility according to which (1) the ability to do otherwise is part of the explanation for why someone is morally responsible, and (2) the ability to do otherwise is not to be

VII

We have seen that attending to the distinction between necessity and grounding has not only clarified Fischer and Ravizza's view but has also provided a clearer view of how it differs from rival actual-sequence approaches as well as from alternative-possibilities approaches. In conclusion, let us briefly return to Cohen's argument against Fischer and Ravizza from the possibility of time travel. We are now in a better position to appreciate why it seemed appealing in the first place, despite its unsoundness.

Recall Cohen's story: young Zoe responds to older Zoe in self-defense, and there is no world in which Zoe refrains from acting in self-defense since older Zoe's existence depends counterfactually on young Zoe's responding in self-defense. Cohen takes this case to raise a problem for Fischer and Ravizza since young Zoe seems not to be responsive to reasons, on their account, and yet an intrinsic duplicate of young Zoe could be responsive to reasons in different circumstances (where the self-defense is in response to someone whose existence does not depend counterfactually on Zoe's response). Crucially, the problem is that there do not seem to be any differences in the grounds of young Zoe's moral responsibility from one case to the next, despite the difference in facts about their alternative possibilities. In other words, the case of time travel that features in Cohen's objection to Fischer and Ravizza allows us to falsify counterfactuals about young Zoe without altering any of the actual-sequence facts about young Zoe's moral competence that ground her moral responsibility.

We have argued that Cohen's argument is unsound, but there is an important lesson to learn from the argument nevertheless, which is that actual-sequence compatibilists ought to de-emphasize, or at least properly contextualize, the role that counterfactuals play in their theories. To the extent that it seems like those counterfactuals are doing the work of *grounding* an agent's moral responsibility, the theory will seem vulnerable to the sort of objection that Cohen launches. Whatever reasons-responsiveness is, it needs to be conceived

---

analyzed in terms of counterfactuals, but instead is to be taken as more fundamental than the counterfactuals it supports. This sort of theory would resemble Fischer and Ravizza's in that moral responsibility is fully explained by facts about the actual sequence, yet it would differ from Fischer and Ravizza's in appealing to an ability to do otherwise. Fischer and Ravizza are interested in distancing themselves from those two sorts of theorists: those who think the ability to do otherwise is required for moral responsibility, and also those who think that facts about other worlds are part of what grounds moral responsibility. What we are pointing out here is that those two sets of theorists are disjoint.

as something that generates its associated counterfactuals rather than being constituted or constrained by them.[42]

Samford University
taylor.w.cyr@gmail.com

Western Washington University
tognazn@wwu.edu

REFERENCES

Arntzenius, Frank, and Tim Maudlin. "Time Travel and Modern Physics." In *Time, Reality and Experience*, edited by Craig Callender, 169–200. Cambridge: Cambridge University Press, 2002.

Clark, Michael, and Liggins, David. "Recent Work on Grounding." *Analysis* 72, no. 4 (October 2012): 812–23.

Clarke, Randolph. "Dispositions, Abilities to Act, and Free Will: The New Dispositionalism." *Mind* 118, no. 470 (April 2009): 323–51.

Cohen, Yishai. "Reasons-Responsiveness and Time Travel." *Journal of Ethics and Social Philosophy* 8, no. 3 (January 2015): 1–7.

Correia, Fabrice. "Ontological Dependence." *Philosophy Compass* 3, no. 5 (September 2008): 1013–32.

Cyr, Taylor W. "Semicompatibilism: No Ability to Do Otherwise Required." *Philosophical Explorations* 20, no. 3 (2017): 308–21.

Dowe, Phil. "The Case for Time Travel." *Philosophy* 75, no. 293 (July 2000): 441–51.

Fara, Michael. "Masked Abilities and Compatibilism." *Mind* 117, no. 468 (October 2008): 844–65.

Fine, Kit. "Ontological Dependence." *Proceedings of the Aristotelian Society* 95, no. 2 (June 1995): 269–90.

Fischer, John Martin. *Deep Control: Essays on Free Will and Value*. New York: Oxford University Press, 2012

———. "The Freedom Required for Moral Responsibility." In *Virtue, Happiness, Knowledge: Themes from the Work of Gail Fine and Terence Irwin*, edited by David O. Brink, Susan Sauve Meyer, and Christopher Shields, 216–33. New York: Oxford University Press, 2018.

42 Thanks very much to Ryan Wasserman and two anonymous reviewers for helpful comments on a previous version of this paper.

———. *My Way: Essays on Moral Responsibility*. New York: Oxford University Press, 2006.

Fischer, John Martin, and Mark Ravizza. *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press, 1998.

Frankfurt, Harry. "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy* 66, no. 23 (December 1969): 829–39.

———. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68, no. 1 (January 1971): 5–20.

———. "Reply to John Martin Fischer." In *The Contours of Agency: Essays on Themes from Harry Frankfurt*, edited by Sarah Buss and Lee Overton, 27–32. Cambridge, MA: MIT Press, 2002.

———. "Some Thoughts concerning PAP." In *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, edited by David Widerker and Michael McKenna, 339–46. Aldershot: Ashgate, 2003.

Franklin, Christopher Evan. "Everyone Thinks That an Ability to Do Otherwise Is Necessary for Free Will and Moral Responsibility." *Philosophical Studies* 172, no. 8 (August 2015): 2091–107.

———. "Masks, Abilities, and Opportunities: Why the New Dispositionalism Cannot Succeed." *The Modern Schoolman* 88, nos. 1/2 (January/April 2011): 89–103.

Kittle, Simon. "Does Everyone Think the Ability to Do Otherwise Is Necessary for Free Will and Moral Responsibility?" *Philosophia* 47, no. 4 (September 2019): 1177–83.

Leon, Felipe, and Neal A. Tognazzini. "Why Frankfurt-Examples Don't Need to Succeed to Succeed." *Philosophy and Phenomenological Research* 80, no. 3 (May 2010): 551–65.

Lewis, David. "The Paradoxes of Time Travel." *American Philosophical Quarterly* 13, no. 2 (April 1976): 145–52.

McCormick, Kelly. "A Dilemma for Morally Responsible Time Travelers." *Philosophical Studies* 174, no. 2 (February 2017): 379–89.

Mele, Alfred R. "Agents' Abilities." *Noûs* 37, no. 3 (September 2003): 447–70.

———. "Reactive Attitudes, Reactivity, and Omissions." *Philosophy and Phenomenological Research* 61, no. 2 (September 2000): 447–52.

Sartorio, Carolina. *Causation and Free Will*. New York: Oxford University Press, 2016.

Spencer, Joshua. "What Time Travelers Cannot Not Do (but Are Responsible for Anyway)." *Philosophical Studies* 166, no. 1 (October 2013): 149–62.

Todd, Patrick, and Tognazzini, Neal A. "A Problem for Guidance Control." *Philosophical Quarterly* 58, no. 233 (October 2008): 685–92.

Vihvelin, Kadri. *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*.

New York: Oxford University Press, 2013.
———. "Free Will Demystified: A Dispositional Account." *Philosophical Topics* 32, nos. 1/2 (Spring/Fall 2004): 427–50.
———. "What Time Travelers Cannot Do." *Philosophical Studies* 81, nos. 2–3 (March 1996): 315–30.
Wasserman, Ryan. *Paradoxes of Time Travel*. New York: Oxford University Press, 2018.
Watson, Gary. "Free Agency." *Journal of Philosophy* 72, no. 8 (April 1975): 205–20.
———. "Reasons and Responsibility." *Ethics* 111, no. 2 (January 2001): 374–94.
Whittle, Ann. "Dispositional Abilities." *Philosophers' Imprint* 10, no. 12 (December 2010): 1–23.

Journal of Ethics and Social Philosophy
VOL. 26, NO. 3 · FEBRUARY 2024
https://doi.org/10.26556/jesp.v26i3.2741

PATERNALISM AND EXCLUSION

Kyle van Oosterum

Some philosophers believe that the distinctive wrong of paternalism has to do with taking a paternalizee's well-being as a *reason* for one's action.[1] This belief serves as a starting point for what I will call the *exclusionary strategy* for determining the wrongness of paternalism. The exclusionary strategy aims to show how some feature of the paternalizee's normative situation morally excludes acting for the paternalizee's well-being or benefit. In this paper, I explain what is wrong with the exclusionary strategy and offer an alternative "nonexclusionary" approach.

Before proceeding, I wish to highlight (and perhaps disappoint some readers in the process) that I will pay comparatively little attention to what *paternalism* means. That question merits its own paper and indeed has generated its own literature.[2] That being said, it will be helpful to have a rough idea of the phenomenon I have in mind. A useful starting point might be Gerald Dworkin's three *jointly sufficient* conditions for paternalistic intervention:

1. *Interference Condition*: An act *Z* (or its omission) interferes with the liberty or autonomy of *Y* (the paternalizee).
2. *No-Consent Condition*: *X* (the paternalizer) does so without the consent of *Y*.
3. *Improvement Condition*: *X* does so just because doing *Z* will improve the welfare of *Y* (where this includes preventing his welfare from diminishing) or in some way promote the interests, values, or good of *Y*.[3]

1 Groll, "Paternalism, Respect, and the Will"; Enoch, "What's Wrong with Paternalism"; and Parry, "Defensive Harm, Consent, and Intervention" and "What's Wrong with Paternalism?"

2 Kleinig, *Paternalism*; Feinberg, *Harm to Self*; Coons and Weber, "Introduction"; Dworkin, "Defining Paternalism"; and Bullock, "A Normatively Neutral Definition of Paternalism."

3 Dworkin, "Paternalism" and "Defining Paternalism."

All of these conditions have been criticized in some way, and other characterizations of paternalism offer useful refinements.[4] Nevertheless, these conditions—or some suitably refined version of them—are often invoked not just in lay conversations about paternalistic policymaking but also in moral debates on the oft-assumed wrongness of paternalism. Consider an example where all three conditions come into play:

> *Fried Chicken*: Frida is a normal adult who wants to eat delicious but unhealthy fried chicken. Her local government, being motivated by a concern for the physical well-being of its constituents, has decided to implement a tax on fried foods.

On Dworkin's account, Frida's government has discouraged her consumption of fried foods and, in so doing, interfered with her liberty, without her express consent, to improve her well-being. This seems like an instance of paternalistic intervention. Note, however, that this definition does not accommodate the assumption that paternalism is presumptively morally wrong. Indeed, in the example above, it might not be crystal clear whether the government in question has acted wrongly. This reflects Dworkin's assumption that we should generally prefer normatively *neutral* definitions and not smuggle in evaluative judgments about the concept we are defining unless, by not including those judgments, we fail to represent it adequately.[5]

Now, I will not take a position on Dworkin's methodological assumption, but the third condition in his definition of paternalism will be essential for what follows. This is because the group of philosophers initially mentioned believe that the distinctive wrong of paternalism has something to do with the "because" part of that improvement condition. If these philosophers are right, they will have vindicated the idea that part of our concept of "paternalism" consists of its pointing to a presumptively problematic practice.

The structure of my paper is as follows. In section 1, I spell out the details of the exclusionary strategy and its motivations. To set up my critique, I distinguish between two versions of the exclusionary strategy by borrowing from the literature on exclusionary reasons. The appeal to second-order exclusionary reasons (i.e., reasons not to act on our first-order reasons) offers a good way of characterizing views that fall under the exclusionary strategy. Section 2 tackles the "justificatory" version of the exclusionary strategy before turning to the "motivational" version. After examining several problems for how to

4 Shiffrin, "Paternalism, Unconscionability Doctrine, and Accommodation"; Grill, "The Normative Core of Paternalism" and "Antipaternalism as a Filter on Reasons"; and Groll, "Medical Paternalism."

5 Dworkin, "Paternalism."

develop these views plausibly, I turn in section 3 to a brief sketch of an alternative approach to determining the wrongness of paternalism. I argue that my nonexclusionary approach is a better way of obtaining the appealing aspects of the exclusionary strategy and cohering with the mainstream view that paternalism is *pro tanto* wrong. As Christian Coons and Michael Weber put it:

> Normative debates about paternalism . . . don't usually concern *whether* it is problematic but *how* problematic it is. . . . There is always at least some *pro tanto* reason to avoid it.[6]

In this paper, I accept that paternalism is *pro tanto* wrong as this view is assumed (sometimes explicitly) by proponents of the exclusionary strategy. Of course, a good philosophical argument may convince us that paternalism is never permissible, but the justificatory bar for this will be high. Nevertheless, I show that both versions of the exclusionary strategy are inconsistent with this mainstream view, *contra* what its defenders claim. This is a surprising result that again motivates consideration of an alternative view that can accommodate the mainstream view. Correspondingly, this paper argues that the exclusionary strategy is problematic while suggesting a more familiar route for determining what makes paternalism wrong. Construing our normative reasons against paternalistic intervention in an exclusionary, second-order way creates many of the problems I cover in section 2. Instead, I argue that we have *first-order* reasons for and against intervention and that their weights can be discerned and balanced against one another to determine the wrongness of paternalism. As such, an overarching aim of my paper is to show that appeals to exclusionary reasons generate implausible implications and are unnecessary in debates concerning the (*pro tanto*) wrongness of paternalism.

1. THE EXCLUSIONARY STRATEGY

The exclusionary strategy, as I have called it, has been defended explicitly by at least three philosophers.[7] These philosophers differ subtly in how they motivate and conceptualize the exclusionary strategy, but they can be grouped roughly into two subcategories. Borrowing from the literature on exclusionary reasons, we can say that there are *motivational* and *justificatory* interpretations of exclusion.[8] Though not every proponent of the exclusionary strategy uses

6 Coons and Weber, "Introduction," 2–4.

7 Groll, "Paternalism, Respect, and the Will"; Enoch, "What's Wrong with Paternalism"; and Parry, "Defensive Harm, Consent, and Intervention" and "What's Wrong with Paternalism?"

8 Adams, "In Defense of Exclusionary Reasons."

the same terminology, the mechanisms described are essentially that of Razian exclusionary reasons.[9] Where David Enoch and Daniel Groll defend a motivational account of exclusion, Jonathan Parry defends a justificatory account of exclusion to explain the wrongness of paternalism. I will outline these two views of exclusion before turning to specific problems in the next section.

1.1. The Motivational Account

The motivational account of the exclusionary strategy focuses on the reasons *for which* an agent may act. Enoch and Groll both appeal to the idea of exclusionary reasons for action. An exclusionary reason is a reason *not* to act on some reason; it defeats or "excludes" a *first-order* reason to do some action but does not outweigh it.[10] Let us take Joseph Raz's "Ann the Banker" example to see how these types of reasons function in everyday deliberation about what we should do. Ann is a banker who, exhausted after a long day of work, nevertheless has to make an important decision about some financial deal. The fact that she is exhausted seems to give her a reason *not* to act on her best judgment of the reasons for and against making this important investment. In Raz-speak, Ann has an exclusionary reason.

With more of a grip on the concept of an exclusionary reason, I will introduce the context behind its specific application in these debates. One common thread among liberal or antipaternalist philosophers is their assertion that the *motive* behind paternalistic intervention is essentially *insulting* to the paternalizee, or potential target of our intervention.[11] It is not hard to see why they might think this. When a paternalist is motivated in this way, they believe they know what is best for a person, perhaps better than that person does themselves. This seems problematic insofar as it lines up with another popular liberal idea, which is that the individual essentially knows what is best for them. It is not in anyone else's moral jurisdiction, if you will, to interfere with their choices unless they harm other people.[12] Perhaps then, if there is something wrong with paternalistic intervention, it resides in the *negative* beliefs and judgments we have about people's choices and whether those are good for them to make.

However, there will be cases where a potential paternalizee does not know what is good for them. Enoch believes (and I agree) that there is nothing wrong with simply having a true belief about whether a paternalizee's actions will

9 Raz, *Practical Reason and Norms*.

10 Clarke, "Exclusionary Reasons"; and Raz, *Practical Reason and Norms*.

11 Feinberg, *Harm to Self*; Shiffrin, "Paternalism, Unconscionability Doctrine, and Accommodation"; Quong, *Liberalism without Perfection*; Begon, "Paternalism"; Cholbi, "Paternalism and Our Rational Powers."

12 Mill, *On Liberty*.

diminish their own well-being. If the paternalist knows that the paternalizee's choices will cause the paternalizee harm, what could be wrong with simply holding that belief? Enoch's suggestion is that the wrongness may consist of a paternalist being motivated to *act* on this belief about the paternalizee's choices.

This is where exclusionary reasons come into the picture. Let us take Enoch's example, which he borrows from Jonathan Quong. Your friend wants to borrow money that you are sure he will use to make himself worse off (perhaps by buying drugs). If you simply believe that he will use the money in a bad way, and you are probably right, there does not seem to be anything wrong with that. Where the wrong lies, Enoch argues, is in acting on that belief and ignoring your friend's questioning ("What's it to you what I do with this money?"), because in so doing you deny the value of your friend's autonomy over their own life.[13] If there is something wrong with paternalizing here, it is because your friend's autonomy gives you an exclusionary reason *not* to act for the reason that it would be good for their well-being if you did not give them the money.

In a similar vein, Groll takes paternalism to be wrong because of how it treats the *will* of the potentially paternalized individual. Roughly, the idea is that a paternalizee's will is intended to silence, trump, or exclude the "reason-giving force" of the other considerations that might be at play when one (a potential paternalizer) is practically deliberating about what to do on behalf of the paternalizee.[14] Groll imagines a medical scenario where a doctor performs some operation and considers a patient's wish not to have the operation as an ingredient in her deliberation about what would be good for the patient's well-being. Groll points out that the patient might be annoyed with the doctor's construal of their will as part of *her* deliberation and not itself the *decisive* factor about whether or not to perform the operation. In other words, as Groll puts it, the patient's will should have made "irrelevant" questions about whether it is good for them to have such an operation.

On both of these accounts, the thought is that a potential paternalizer acts wrongly in being motivated solely (or overridingly) by considerations of a paternalizee's good. They hold that the paternalizee's autonomy or will morally excludes such considerations as reasons for action. Importantly, neither Groll nor Enoch believes that paternalism is always wrong, and each has suggested

13 It is unclear whether Enoch has his own specific conception of autonomy in mind. See Enoch, "Hypothetical Consent and the Value(s) of Autonomy." For our purposes, we can interpret it broadly as a person's ability to make decisions in line with their values or conception of the good life. See Birks, "How Wrong Is Paternalism?"

14 Groll, "Paternalism, Respect, and the Will," 701.

that his is an account of the *pro tanto* wrongness of paternalism.[15] Each maintains that paternalism is usually but not always wrong and believes that his account of the exclusionary strategy can vindicate that judgment. Recall that such a view is the mainstream one in philosophical writing about paternalism. It would be interesting if it turned out that their account was not consistent with this view (more on that in section 2.2).

1.2. The Justificatory Account

Let us turn now to the justificatory account. Parry's views on the wrongness of paternalism exist in the larger context of defensive harm, but I believe they fit well under the banner of the exclusionary strategy. Like Enoch and Groll, Parry is trying to figure out why it can be wrong to (paternalistically) act for someone's good or well-being. His response to this question appeals to the idea of a *moral power*, that is, the ability persons possess to change the moral or normative landscape around them (e.g., by changing what it is permissible to do to them).

For example, when a person consents to sexual intercourse, they make what is usually impermissible—another person interfering with their bodily integrity—into something permissible. Parry believes that just as we have the power to control our bodies and property (our material resources), we also have the power to control the use of our "good," where "good" refers to reasons grounded in our well-being (our "normative resources").[16] To use someone's good, he claims, is to justify one's actions by appealing to the fact that it would be good for them if we did that. Let us return to the example offered by Quong above. Your friend has the power to make their good "inadmissible" as a justifying reason for action, such that declining to give them the money cannot be justifiable (for the reason that it would be good for them).

Notice that talk of the inadmissibility of a reason sounds very similar to the exclusionary reasons mentioned by Groll and Enoch. To my mind, it sounds similar because Parry is defending a justificatory account of exclusion. A justificatory account of exclusion holds that exclusionary reasons essentially change the "right-making" features of an action; they exclude or prevent ordinary moral reasons from standing in a justifying relation to actions.[17] Let us consider a nonpaternalistic example of this phenomenon. Adams argues that laws can be thought of as (exclusionary) reasons that exclude reasons that might count

15 Groll's recent views on the wrongness of paternalism seem to involve much more rights talk than talk of exclusionary reasons. See Groll, "Paternalism and Rights."

16 Parry, "What's Wrong with Paternalism?"

17 Moore, "Authority, Law and Razian Reasons"; and Adams, "In Defense of Exclusionary Reasons."

in favor of law-breaking, such as pulling over on a highway to help a wounded animal.[18] A distinguishing feature of exclusionary reasons is that they do not compete in weight with first-order reasons and generally have absolute priority over the reasons that they exclude.[19] But if this is true, then even though we could have incredibly weighty reasons to help the animal, the law makes those seemingly weighty reasons play *no* justificatory role whatsoever in our deliberation. To my mind, the same thing is going on in Parry's account of the wrongness of paternalism. As he puts it, reasons to promote a person's well-being become "unavailable" as justifications for action by virtue of an exercise of our moral power (to exclude the use of our "good").[20]

At this point, it might be helpful to distinguish between motivational and justificatory exclusionary strategies. The motivational account locates the wrong of paternalism in the well-being-related reasons that a paternalizer chooses to act on. Autonomy (or the will, in Groll's account) provides an exclusionary moral reason for the paternalizer not to act for the good of the paternalizee. The justificatory account makes no reference to a paternalizer's motivations for action. Instead, it focuses on how features of the situation make well-being-related reasons the wrong sort of reason to act on. This is because they are no longer part of the potential right-making reasons for justifying action. Whereas the justificatory account denies the ordinary justificatory role that well-being-related reasons play, the motivational account does not make this claim about reasons. Well-being-related reasons exist as strong reasons to act on, but it so happens that respect for autonomy or the will makes it so that such reasons are wrong to be motivated by. In short, the wrong lies either (i) in acting on a reason that no longer performs its function (the justificatory account) or (ii) in acting on a wrong yet functional moral reason (the motivational account).

2. Problems for the Exclusionary Strategy

In this section, I will argue that both versions of the exclusionary strategy are problematic. I will show that both views struggle to accommodate the mainstream view of paternalism's *pro tanto* wrongness that also counts against them. Upon close examination, the justificatory account, while clearly specifying how well-being is to be excluded, delivers counterintuitively strong verdicts that seem never to countenance paternalistic intervention (when it seems

18 Adams, "In Defense of Exclusionary Reasons."

19 Raz, *Practical Reason and Norms*.

20 Parry, "What's Wrong with Paternalism?"

permissible). By contrast, the motivational account enjoys some intuitive advantages over the justificatory account, but it is unclear how to specify its exclusion of well-being in a plausible way. In section 3, I offer a general diagnosis of why these views go wrong, as well as an alternative view that outperforms them both. For now, if neither exclusionary account turns out to be plausible, this supports my contention that it is unnecessary and implausible to appeal to exclusionary reasons to explain the *pro tanto* wrongness of paternalism. This is because there may be alternative views, such as my own highlighted in section 3.2, that can vindicate much of the exclusionary strategy's appeal without a second-order level of reasoning and without the problems that such reasoning gives rise to.

2.1. Problems with Justificatory Exclusion

An important caveat to Parry's moral power account is that a person has to be able to *competently* refuse to be benefitted by others. More precisely, a person has to competently exclude the use of their good as a justification for someone's action toward them. This is a principled qualification inspired most likely by the oft-cited distinction between *soft* and *hard* paternalism.[21] Though that distinction has come under fire, the thought is plausible enough: paternalistically interfering with someone seems less wrong if a person made their choice involuntarily. This involuntariness could be due to the individual not being an adult yet, being under the influence of alcohol or drugs, or perhaps suffering from some physical or mental ailment. Roughly, soft paternalists believe that whether a person's choice is voluntary is relevant to the justifiability of paternalistic intervention. By contrast, hard paternalists disagree that voluntariness should always matter.[22] In practice, correct judgments of voluntariness can be hard to make, but it is *prima facie* plausible to include them as features that help justify paternalistic intervention. In any case, if a person "incompetently" refused a benefit, then this would lead to the intuitive verdict that we could still use their good as a justification for paternalistic action (assuming that such an action would count as "paternalistic" in the first place).

So far, so good. However, we might think cases of incompetent refusal are the low-hanging philosophical fruit for this debate. After all, some philosophers do not regard soft paternalism as a kind of paternalism at all.[23] The real challenge to Parry's justificatory account would be to identify one case where

21 Feinberg, *Harm to Self*.

22 This is an obvious caricature of a sophisticated debate that I am mentioning only to provide context for what follows. For evaluation of the distinction between soft and hard paternalism, see Hanna, "Hard and Soft Paternalism."

23 Feinberg, *Harm to Self*.

a competent refusal has occurred yet a paternalistic intervention would not be wrong. Consider an adapted version of Richard Arneson's famous case:

> *Pouting Young Adult*: Tom is unreasonably distressed at some disappointment he has suffered. Perhaps he has been bested in competition for a job he coveted.... Perhaps a particularly charming rabbit he saw at the Humane Society pet adoption center and hoped to choose and make his pet was adopted by another person. Whatever the cause of his distress, he is unhappy, feels vaguely cheated by the world at large, and wants at the moment nothing more than to express his disappointment by committing suicide. In addition, Tom knows he will likely change his mind but right now has no interest in doing so. He is neither mentally ill nor incompetent as a decision-maker. He simply wants to commit suicide and has refused appeals by his friends to change his mind and think of his own well-being.[24]

To my mind, this is a case where paternalistic intervention seems not only permissible but justified. Of course, a very staunch antipaternalist might just deny that it is intuitively permissible to interfere here. However, it is hard to see how if paternalism were not permitted here it would still be permitted in a similarly extreme case. It seems that the justificatory account, with its notion of a "competent refusal," makes such a paternalistic action unjustifiable. This may lead us to wonder how paternalism toward competent adults could ever be permitted on this account.[25]

But this is far too quick. Proponents of this justificatory account might appeal to the distinction, captured nicely by David Owens, between *acting wrongly* and *wronging someone*.[26] Another way of putting this is that we might think we can commit a wrong without doing the wrong thing. When I break a promise to meet my friend to help another person who has been hit by a car, I have wronged my friend but not done the all-things-considered wrong thing. Here, my promise-breaking is intuitively justified, which suggests, as Owens puts it, that "committing a wrong can be the right thing to do."[27] This idea fits in well with "exclusionary reasons" terminology, because one can think of a promise as excluding the reasons *not to act* on or break the promise.

24 Arneson, "Joel Feinberg and the Justification of Hard Paternalism," 278–79.

25 This point has also been noted by Quong in his recent talk on Parry's account of antipaternalism, "Paternalism, Disagreement and Groups."

26 I thank Lorenzo Elijah again for pointing this distinction out to me.

27 Owens, *Shaping the Normative Landscape*, 45.

Perhaps then, in cases such as Pouting Young Adult, advocates of the exclusionary strategy can claim that although we have wronged the paternalizee, we have not acted wrongly in paternalistically interfering (in those extreme cases). I have not attended to this fact: just because someone has been wronged does not mean that what has been done is wrong or impermissible. Exclusionary strategy proponents can claim that permissible paternalistic intervention involves cases of "permissible wronging," so to speak, which allows them to maintain that paternalism is wrong but not always wrong. In other words, not all paternalistic wrongings are wrongs. This seems like a plausible enough conclusion to hold.

Unfortunately, this appeal is unavailable to proponents of the justificatory account of the exclusionary strategy. Parry's view renders well-being-related reasons counting in favor of paternalistic intervention disabled or unable to play any justificatory role for action.[28] Obviously, according to this view, if a paternalizer were to intervene on the basis of well-being-related reasons, they would naturally wrong the paternalizee. But what makes a paternalistic intervention in Pouting Young Adult "not" wrong? One might think the intervention is intuitively permissible and all-things-considered justified, but the *content* of this intuition and justification is surely the very same well-being-related reason that is disabled by exclusion. If some other non-well-being-related reason forms the intuitive justification for intervention, then we are not plausibly dealing with a case of paternalism anymore. After all, the exclusionary strategy's account of paternalism relies on the notion that the justification for the intervention is well-being-related (see introduction).

So, we have something of a dilemma. Adherents of the justificatory account cannot defend the idea that a paternalistic wronging would not be wrong. They cannot appeal to well-being-related reasons, and they need those very reasons to be discussing a "paternalistic" act in the first place. In other words, either they must accept that every paternalistic wronging is indeed wrong—an extreme conclusion—or the act of intervention is "not wrong" but can no longer be described as "paternalistic." Therefore, this Owens-style idea cannot be used to square the justificatory account with the *pro tanto* view of paternalism's

28 Parry has suggested to me that there could be a positive and a negative way to read his view. On the negative reading, his view states that welfarist reasons are not there to justify the action. On the positive reading, the use of the paternalizee's welfarist reasons is just a directed wrong to the paternalizee (e.g., a form of trespass). Perhaps a version of his view could be developed with only the positive reading. There are two problems here. First, the negative reading contributes to making it an "exclusionary" view in the first place. Second, and related, one might wonder how distinctive his view would be from other antipaternalist views without this negative claim.

wrongness. The result is that this view, while clear in its formulation, is a counterintuitively strong version of antipaternalism and cannot make room for intuitively permissible cases of paternalism.

Here is another concern with Parry's view. Recall Parry's analogy between material resources and normative resources. The inference drawn from this analogy is that "wrongable" paternalizees determine the moral status of paternalistic intervention because moral reasons *belong* to them. The idea that our reasons "belong" to us is mysterious. I think there is an importantly relevant distinction between claiming that these reasons are *about the paternalizee* and saying that these reasons *are theirs*. The former claim is straightforward and makes sense. After all, some philosophers think that it is wrong to act or be motivated by the reasons that *refer* to a paternalizee's well-being (e.g., what the motivational account of exclusion seeks to defend). The latter claim, namely, that moral reasons (i.e., reasons having to do with well-being) can be ours to control, strikes me as implausible and in need of further defense. Obviously, this taps into a deeper question about whether reasons can be "up to us" in a metaphysical sense that is admittedly not Parry's focus.[29] While Parry does offer a number of rationales in favor of having a moral power to exclude reasons, he has not shown that we have this power; in other words, it is still unclear how these reasons are (or become) *ours* in the way that material property is ours.[30] For now, this contestable analogy seems to be justifying the existence of this power and our supposed ownership of reasons. Therefore, Parry's argument is not only implausible as an account of paternalism's *pro tanto* wrongness, it seems also to be derived from implausible footings.[31]

2.2. Problems for Motivational Exclusion

2.2.1. A Prima Facie Problem and the Scope of Exclusionary Reasons

I want to suggest that the following insight can be gained from the justificatory version of exclusion: claiming that well-being-related reasons do not feature at all in a moral assessment of paternalistic intervention is unnecessary. It is unnecessary with respect to reaching the conclusion that paternalism is *pro tanto* morally wrong. Indeed, as I have just argued, the justificatory version of exclusion makes it difficult to render any paternalistic intervention permissible.

29 Moore, "Authority, Law and Razian Reasons"; and Chang, "Do We Have Normative Powers?"

30 I believe the strategy Parry pursues is to justify the power in virtue of how it serves the realization of some important value. However, this does not show that the power exists, nor does it dispel the mysterious claims about the ownership of reasons it seems to involve.

31 I thank an anonymous referee for the suggestion to elaborate the point in this way.

It also makes strange claims about our supposed ownership of moral reasons. Instead, advocates of the exclusionary idea could appeal to the motivational version of the exclusionary strategy (MES), which makes neither of those claims. The MES just argues that to be motivated to act on well-being-related reasons is *pro tanto* wrong.

Why would it be wrong to be motivated by these reasons? Part of what it is to respect autonomy (or treat one's will as structurally decisive, in "Groll-speak" now) morally excludes being motivated by what is good for the paternalizee's life. Importantly, exclusionary reasons, in this sense, are reasons for not being motivated in one's actions by certain "valid considerations."[32] What seems more intuitive about this account than Parry's is that we are not making the extreme claim that well-being is *not* a valid reason-generating consideration and that it *could not* be part of the justificatory story. Instead, the thought is just that the importance of autonomy overrides or generally takes priority over well-being. The device of an exclusionary reason is one way of articulating that thought. This is how we get to the view that autonomy generates an (exclusionary) reason *not* to act on the reason that it would be good for the paternalizee's well-being to interfere.

But does this view do better in cohering with the verdict that paternalism is only *pro tanto* wrong? Enoch and Groll seem to think so, but I believe there are some ambiguities in their account that make this question difficult to answer affirmatively. The chief ambiguity consists in how *much* this exclusionary reason excludes. At the moment, the view looks like this:

First Pass:

P1: Paternalistic interferences are wrong if there are unexcluded moral reasons that favor not paternalistically interfering.

P2: There is an exclusionary reason that is grounded in the paternalizee's autonomy or will. It is an unexcluded moral reason not to interfere for the reason that it would be good for the well-being of the paternalizee (to interfere).

C: Therefore, paternalistic interferences are wrong.

Of course, this statement of the view is far too general. Without qualification, it would rule out any case of paternalistic intervention (targeted at competent adults). This is because exclusionary reasons are generally thought to have *absolute* priority over the reasons that they exclude.[33] At first glance, this argument holds that autonomy (or the will), being the ground of an exclusionary reason,

32 Raz, *Practical Reason and Norms*, 185.

33 Raz, *Practical Reason and Norms*.

always has priority over our first-order reasons to promote one's well-being, regardless of this reason's normative strength. Even very staunch antipaternalists will concede that this is a counterintuitively strong conclusion, which is why the widely held view is that paternalism is only *pro tanto* wrong. Now we can return to the question of whether Enoch and Groll's view is actually consistent with this widely held view despite the conclusion of this first-pass argument.

If there is a problem with the first-pass argument, it resides in P2, which is where some qualifications might be attempted. Perhaps P2 can instead read:

> P2*: There is an exclusionary reason grounded in the paternalizee's autonomy or will. It provides an unexcluded moral reason not to interfere that is *usually undefeated* by the reason that it would be good for the well-being of the paternalizee (to interfere).

P2* would allow us to say that there can be some cases where the exclusionary reason can be outweighed or defeated by the reason to act for the well-being of the paternalizee. This would seem to get the motivational account closer to the widely held view, but it unfortunately comes at the cost of distinctiveness. As Raz himself points out:

> If [exclusionary reasons] have to compete in weight with the excluded reasons, they will only exclude reasons which they outweigh, and thus lose distinctiveness.[34]

The problem with the P2* move is that we lose what makes an exclusionary reason "exclusionary." Exclusionary reasons are reasons that refer to the balance of first-order reasons for performing some action and are *not* supposed to be *part of* that same balance of reasons. In other words, we would simply be saying that autonomy generates a first-order reason not to interfere that is often, but not always, stronger than the first-order reasons well-being gives us to interfere. However, this statement would not be consistent with the motivational account's commitments to the notion of exclusion.

In short, this view seems to fall prey to a dilemma. On the one hand, P2 gives us a consistent statement of this view, but it generates the counterintuitively strong conclusion of the first-pass argument. On the other hand, P2* allows these theorists to avoid this conclusion at the cost of a less distinctive view, which no longer seems exclusionary. Clearly, this view's proponents would not go for either horn of the dilemma. They believe they can coherently defend the

34 Raz, *Practical Reason and Norms*, 189.

view that paternalism is *pro tanto* morally wrong with the device of exclusionary reasons. How would they go about avoiding this dilemma?

An important feature that has been underspecified in the motivational account is precisely what *scope* such exclusionary reasons have—or should have, for that matter. What it means for exclusionary reasons to vary in scope is to say that they might exclude all or only some of the reasons that apply to some situation in practical reasoning.[35] For example, consider Raz's character Colin, who makes a promise to his wife to decide what to do about their son's education only on the basis of their son's interests. Here, Colin has an exclusionary reason not to act on reasons unrelated to his son's interests. However, the scope of that reason does not extend so far as to exclude considerations of justice to other people. Raz's notion of exclusionary reasons is complicated by, but also more faithful to, the circumstantial nature of practical reasoning because of these *scope-affecting* considerations. Indeed, the complication for practical reasoners consists in determining when these considerations narrow the scope of exclusionary reasons such that they no longer exclude conflicting first-order reasons.

How does this bear on the debate about the wrongness of paternalism? Recall that proponents of the MES only want to defend the *pro tanto* wrongness of paternalism. They may want to accommodate cases where a paternalistic intervention is intuitively permissible, such as Pouting Young Adult.

Pouting Young Adult was a dramatic case chosen to elicit the commonsense intuition that it is *prima facie* permissible to interfere with Tom's autonomous choice. Let us translate the details of the case into the MES framework as follows: Tom's autonomy (or will) generates an exclusionary reason not to act on the first-order reason (that is, that it would be good for his well-being if we prevented his suicide). Now, if we assume that MES proponents want to allow for paternalistic interference in this kind of case, what would they have to say? They could appeal to considerations that *affect* the ordinary scope of exclusionary reasons generated by a paternalizee's autonomy or will. Perhaps the scope of autonomy's exclusionary force might be limited to a paternalizee's non-self-annihilating decisions. So, while autonomy excludes acting for the reason that it would be good for a paternalizee's well-being, perhaps it does not exclude a first-order reason to prevent suicide.

However, reining in the scope of the exclusionary reason in this way is somewhat *ad hoc*, and it forces the MES proponent to unnecessarily defend a general prohibition against suicide. I believe that what is lurking in the background is some concern for Tom's well-being and a belief that it is sometimes permissible to act for such a reason. While we normally want to treat a person's autonomy

35 Raz, *Practical Reason and Norms*, 39.

(or will) as decisive in this exclusionary way, cases such as Pouting Young Adult make us hesitate because so much of Tom's well-being is at stake. However, in the case of your friend asking for money, you might feel more compelled to respect the exclusionary force of his autonomy. I think that the asymmetry between these cases might be explained in this way: a paternalizee's well-being sometimes seems to play the role of an excluded reason and sometimes seems to be unexcluded by their autonomy.[36] But how can this first-order well-being-related reason operate in both of these ways? Is there some principled way to distinguish when this well-being-related reason is plausibly excludable or nonexcludable?

2.2.2. Different Ways to Identify the Scope of the Exclusionary Reason

The answer to those questions depends on what account of well-being we are operating with. However, I am not convinced that applying any account of well-being could yield a nonarbitrary answer to the second of those questions. Let us plug in each of Derek Parfit's three accounts of well-being, one at a time, to see why this is the case.[37] First, objective-list theories claim that there is some list of goods, such as knowledge and friendship, that constitute well-being and make an agent's life good whether or not the agent desires them. This is a crude rendering of this theory, but it suffices for our purposes. Perhaps, using the objective-list theory, the MES proponent might suggest that autonomy excludes some of the goods on the objective list but not others. Those goods that autonomy does not exclude would provide a kind of unexcluded well-being-related reason that helps deal with certain cases of intuitively permissible paternalism.

The problem with this approach is that it will be difficult to determine *which* goods should not be excluded and in *which contexts* this ought to be the case. One general problem for objective-list theories is determining what goods plausibly belong on such a list. Here, we have a similar issue: how do we determine which goods belong on this list *and* how can we create a plausible separation between the excluded and unexcluded well-being reasons to which they give rise? Since the objective-list theory donates its conceptual baggage here, the MES proponent should probably not adopt this as their account of well-being.[38]

Second, we could apply some form of hedonism to this question. Perhaps there is a *threshold* for the amount of pain to be prevented (or pleasure to be obtained) that could draw the line between excluded and unexcluded

36 This thought was suggested to me by Enoch on an earlier draft of this paper.

37 Parfit, *Reasons and Persons*, app. 1.

38 A further issue might be that implementing the objective-list theory conflicts somewhat with the spirit of autonomy's exclusionary scope. It might be strange that certain objective goods that *I* do not think are objectively good play some role in deciding when I am wronged by paternalistic intervention.

well-being-related reasons.[39] In Pouting Young Adult, we could say that Tom's death, being the ultimate loss of well-being, renders this decision unexcluded by his autonomy. Since this pain would surpass some threshold, it would be outside the scope of the autonomy-related exclusionary reason and thus defeat said reason. As a result, we could obtain the verdict that paternalistic intervention in that extreme case would not be wrong. For this to be consistent with the *pro tanto* view of paternalism's wrongness, the threshold would have to be very high. I think that this is certainly more plausible than applying the objective-list theory here.

However, I am skeptical that a threshold approach identifies the right scope-affecting consideration for this exclusionary reason. My first concern is about how high the threshold should actually be. To my mind, the threshold approach seems more intuitively appealing the more ambiguously it is defined. Let us say the threshold was defined by the potential death of a paternalizee. One might think that though this is a concrete specification of the threshold, it seems somewhat arbitrary. Why should excruciating pain not satisfy the threshold? When the threshold is high yet ambiguously defined, this will lead to a lot of disagreement about if and when the threshold applies. Perhaps the MES proponent might reply that this is fine, because it mirrors the real-life complexities of practical reasoning about paternalistic intervention. However, insofar as this approach is used to try to distinguish between excluded and unexcluded well-being reasons, it raises more questions than it was intended to answer.

My second concern is that this hedonistic threshold-based approach might, depending on how we characterize it, start to resemble the "objectivist" tendencies of the objective-list theory. This is because the justification for a well-being threshold does not originate in the paternalizee themselves and seems to imply the view that pain or pleasure is worth avoiding or pursuing whatever else the paternalizee might want. No doubt this can be a plausible point of view, but the point of invoking exclusionary reasons is largely to bring such matters under the normative auspices of the paternalizee. That is, we want to let them determine the amount of pain and pleasure they want to receive over the course of their life. So, externally defining well-being thresholds for exclusionary reasons to apply seems troubling and inconsistent with the motivations for the exclusionary strategy.

Finally, we could try some kind of desire-satisfaction theory of well-being. Now, there are many different variants of this theory, so in principle, there are many ways MES proponents could deploy it. Perhaps, they could claim that there are *certain* desires whose satisfaction is not conducive to promoting well-being and that those desires might not fall within the scope of an

39 I thank Lorenzo Elijah for this way of formulating the point.

exclusionary reason. I think this move is already off the table, as we considered it *ad hoc* to rely on ruling out the desire to commit suicide in Pouting Young Adult as an unexcluded well-being reason. In general, it may appear arbitrary to rule out the satisfaction of certain desires just to obtain the intuitively right verdicts about cases.

Instead, we could rein in the scope of the exclusionary reason not by referring to certain desires but to certain *kinds* of desires. Perhaps uninformed desires would not be excluded by autonomy and thus permit paternalistic intervention, whereas informed desires ought to be excluded. The distinction between excluded and unexcluded well-being reasons could just be based on the distinction between the satisfaction of informed and uninformed desires. Again, I think that Pouting Young Adult shows that even on an informed desire-satisfaction theory of well-being, there seems to be some intuitively permissible well-being reason to act on and be motivated by. Arguably then, this way of identifying which well-being-related reasons are excludable fails as well.[40]

In short, the MES cannot be given an articulation to accommodate the *pro tanto* view of paternalism's wrongness. On three plausible ways one could distinguish between excluded and unexcluded well-being-related reasons, the result was that the approaches were either arbitrary or counterintuitive. Though the motivational version of exclusion did not adopt the extreme approach of ruling out the justifying force of well-being-related reasons (as Parry's account seemed to do), it unfortunately could not neatly accommodate them into its framework.

3. MOVING AWAY FROM EXCLUSION: A SKETCH

3.1. Reflecting on Exclusion

Clearly, the proponents of the exclusionary strategy believe that we need to maintain the standard view that paternalism is often but not always morally wrong. The appeal to the normative exclusion of a paternalizee's well-being was thought to be one way to do this, but I have shown that neither version of the exclusionary strategy can be spelled out easily. There is something wrong with treating exclusion as a *constitutive* feature of the wrongness of paternalism rather than one that may explain paternalism's wrongness in some circumstances. It

40 Parry has suggested to me that we could fix the scope of exclusion in a simpler way without discussing different conceptions of well-being. For example, we might think that only a certain *quantity* of well-being can be excluded or only a certain *proportion* of well-being can be excluded. While these would be simpler, it is unclear to me how these views would differ from a threshold account once they are fully elaborated.

simply does not seem like we (always) wrong someone by taking their well-being as a reason for our action (or that we are acting on a reason that no longer plays any justificatory force for action). Another way to put what is going wrong here is to echo Scanlon's observation that invoking exclusionary reasons leads us to ignore the "substantive relevance" of the reasons we are excluding.[41] These are reasons that have to do with a paternalizee's well-being. Such reasons are ordinarily good ones to be motivated by or justifying of action. However, it is possible that in the cases Parry, Groll, and Enoch identify, those reasons are not permissible to act on but perhaps only within a "nonexclusionary" framework.

It is worth stating what the exclusionary view gets right before considering an alternative way of accounting for the wrongness of paternalism. First, the exclusionary strategy can support intuitive verdicts about wrongful paternalism, as in Quong's money-lending case. Second, we might think, as Enoch does, that exclusion generates the correct moral phenomenology associated with paternalism. That is, when paternalizers act, they get involved in what is (morally) not their business, which makes it difficult to justify such actions in a way that is consistent with respecting the other person's autonomy.

So, the exclusionary strategy has these sorts of things going for it. However, the thrust of my paper suggests that going down this route is philosophically costly and onerous. The natural thing to do is to develop an alternative philosophical account—that is, an account that obtains the goods listed above and the verdict that paternalism is *pro tanto* wrong but does without talk of exclusionary reasons and the problems created by the exclusionary strategy. Importantly, this is not to say that we do away with reasons-talk for the wrongness of paternalism, but that we adopt a more familiar approach of reasoning on the *first-order* level. I call this the nonexclusionary approach.

3.2. The Nonexclusionary Approach

The view I have in mind is moderate without conceding too much to a position identified by Jason Hanna as "pro-paternalistic."[42] Like Hanna, I think it is always a valid reason-generating consideration to act in someone's best interest or for the promotion of their well-being. Of course, just because that reason might be valid to act on does not mean that it will be *decisive* in all or even many cases. The idea on the table, then, is that well-being-related reasons (to paternalistically interfere) will normally vary in strength or weight. They will act in competition with reasons to *refrain* from interfering, which might be

41 Scanlon, "Reasons," 241.

42 Hanna, *In Our Best Interest*, 1.

autonomy or will-related (or some other antipaternalistic unit of concern).[43] The idea of balancing our (first-order) reasons for and against paternalistically interfering is not unfamiliar to the literature. However, remarkably little has been said about how to discern the strength or weights of these reasons. To that end, I think it would be helpful to turn to another idea in the literature on practical reasoning: *modifiers*.

Modifiers are facts that, though not themselves reasons, are capable of directly affecting the weight of a reason for action.[44] For example, imagine you have a desire to eat Kentucky Fried Chicken (KFC) which can plausibly give you a reason to go eat some KFC right now. However, the fact that it is rush hour and there will be traffic on the way to KFC might make you less keen to go eat some KFC now. Traffic is not itself a reason to not eat KFC, but it appears to *weaken* your reason to go eat KFC now, given that you do not want to spend so much time in traffic.

Modifiers come in two varieties. The example outlined above displays an *attenuator* in action, a fact that weakens the weight of a reason to do something. By contrast, *intensifiers* are facts that increase the weight of a reason to do something. For example, imagine you are walking around and notice a person who needs help. The fact that this person needs help presumably gives you some reason to help her. But the fact that you are the *only* person around who can help seemingly strengthens your reason to help.[45] That you are the only person around is not itself a reason to help, and the same would be true if you were one of many bystanders. However, that you are the only person around "intensifies" the weight of your reason to help if (and when) this reason exists. In short, modifiers can affect the weight of our reasons and play an important role in helping us to decide what action we should take or are justified in taking.

In the context of paternalism, there might be all sorts of facts that strengthen or weaken both our first-order reasons to paternalistically interfere and our reasons to refrain from interfering. For example, the amount of well-being that could be promoted (or prevented from being diminished) might intensify a reason to interfere. The significance of one's autonomous choice might also modify the strength of a reason to refrain from interfering. Another potential modifier might be the *closeness* of the relationship a paternalizer has to a prospective paternalizee. Perhaps the more intimately related paternalizers are with paternalizees (i.e., paternalism between friends), the stronger a reason

43 This move is currently being considered by other philosophers too, and my sketch gestures at ways in which it can be made more precise. See Shafer-Landau, "Liberalism and Paternalism"; and Birks, "How Wrong Is Paternalism?"

44 Bader, "Conditions, Modifiers, and Holism."

45 Dancy, *Ethics without Principles*.

becomes to paternalistically interfere.[46] There may be many more kinds of modifiers, and much more could be said in defense of these particular ones. I think these are helpful enough heuristics for discerning the strength or weights of these reasons in a variety of cases.

So much for my view. But what do we get when we couch the wrongness of paternalism simply in the terminology of first-order reasons and their modifiers? First, I believe we can already get intuitive verdicts about cases such as Pouting Young Adult. What we have there is a conflicting well-being-related reason to interfere and an autonomy-related reason to refrain from interfering. However, the well-being-related reason seems intensified by the amount of well-being at stake (i.e., the rest of Tom's possibly good life). Conversely, the autonomy-related reason seems attenuated by the fact that, by Tom's own lights, the choice does not seem that significant to him. Ergo, the well-being-related reason defeats the autonomy-related reason, which matches our intuitive verdict about this case being one of permissible paternalism.

Second, though paternalism is permissible here, we can still obtain the verdict that proponents of the exclusionary strategy want, namely, that paternalism is *pro tanto* wrong. In fact, the wrongness can still be tied to well-being. We should not claim that it is wrong to justify one's action on the basis of well-being (or be motivated by such a reason) because such reasons are excluded. Rather, the wrongness consists in acting on a well-being reason that has been *defeated* because it is weaker than a reason to refrain from interfering. To make this view consistent with the thought that paternalism is *pro tanto* wrong, one need only show how such well-being-related reasons might generally be weaker. They can appeal to a variety of the modifiers I suggested above to justify such a judgment. Crucially, we obtain a view of paternalism's *pro tanto* wrongness without the strong and counterintuitive commitments of the exclusionary views. That is, we need not say that these reasons belong to us, that some reasons can be made to have no justifying force, or that it is always wrong to be motivated by a certain class of reasons.

But what about the phenomenological point? Can this nonexclusionary approach still accommodate those strong (but not absolutely strong) antipaternalistic intuitions? One might worry that a paternalizer's determination of reason strengths and balancing of reasons already violates the "not your business" connotations of valuing a person's autonomy.[47] Another way to put it is that there is an important distinction between *recognizing* a conflict of reasons

46 There is an active debate in the paternalism literature about whether there is a morally relevant difference between paternalism that is practiced by one's intimates or by the state. See Tsai, "Paternalism and Intimate Relationships"; and Birks, "Sex, Love, and Paternalism."

47 This is another point made to me by David Enoch on an earlier draft of a paper.

and *imposing* a view as to how this conflict between reasons should be adjudicated.[48] Of course, there is no way to avoid an imposition about how to resolve such reasons conflicts. The antipaternalist in some sense "imposes" their view that reasons to refrain from interfering should generally prevail over reasons to interfere. Unsurprisingly, this is a "welcome" imposition in what is a generally antipaternalistic climate of philosophical writing.

I do not have the space to develop a comprehensive answer to this issue, as it is not the focus of my paper. However, I believe my view can affirm that it is generally not a paternalist's business to interfere, but it can only do so if we are clearer on what it is to value autonomy. We could value autonomy in two ways: either we value its *possession*, or we value its *exercise*.[49] If the latter is so, I do not think the intuition that it is not our business to interfere will always be so strong. In cases such as Pouting Young Adult—or a variant of that case where Tom is a close friend of ours—we might think it utterly callous not to do something and get involved. Perhaps this is because what matters is not the fact that Tom is autonomous but that he is exercising his autonomy in a problematic way.

The idea, then, is that certain exercises of autonomy have more value than others and that it is those valuable exercises of autonomy that make us think it is not our business to interfere. For example, when an unconscious Jehovah's Witness is given a blood transfusion, we might think this is problematic precisely because the expression of a religious belief is a valuable exercise of one's autonomy. Here, we could plausibly think it is not our business to save their life. So, if autonomy's value is linked somehow to its exercise, then we will not always think it is not our business (not to interfere). If that is true, then my view can still map onto this somewhat revised antipaternalistic phenomenological datum.

4. Conclusion

I hope to have shown that the exclusionary strategy is problematic due partly to how difficult it is to elaborate and because it does not square well with the mainstream view of paternalism's *pro tanto* wrongness. Importantly, I do not think that problems with the exclusionary strategy should raise any concerns about the viability of exclusionary reasons in general. It should not do so because the application of exclusionary reasons to any domain of philosophy will come with its own unique intricacies and theoretical baggage. Nevertheless, perhaps because the exclusionary strategy is still being developed, the problems I have

48 Malm, "Feinberg's Anti-Paternalism and the Balancing Strategy," 198.

49 I am borrowing here from Raz's discussion of autonomy. See Raz, *The Morality of Freedom*, 370.

raised may yet be resolvable. In that case, this paper can be read as an invitation to antipaternalists drawn to these ideas to deal with the complexities. It may also be that the exclusionary strategy is so appealing (for other reasons) that these complexities, if not resolvable, might be taken in stride.

That being said, I believe we can salvage the exclusionary strategy's appeal and maintain a similar antipaternalistic stance with a normative toolkit that is more familiar and on a run-of-the-mill first-order level of reasoning. In a way, the idea of balancing reasons for and against paternalistic intervention is a commonsensical one. What I hope to have added to this commonsense view is some more precision by adding modifiers to the debate. We should focus on not only the reasons for and against paternalistically interfering but also what might specifically influence the strength of those reasons. More can be said in defense of the view I have developed, but this sketch is an important step toward using normative reasons in the context of paternalism's wrongness in an intuitively better way.[50]

Hertford College, University of Oxford
kyle.vanoosterum@philosophy.ox.ac.uk

REFERENCES

Adams, N. P. "In Defense of Exclusionary Reasons." *Philosophical Studies* 178, no. 1 (January 2021): 235–53.

Arneson, Richard J. "Joel Feinberg and the Justification of Hard Paternalism." *Legal Theory* 11, no. 3 (September 2005): 259–84.

Bader, Ralf. "Conditions, Modifiers, and Holism." In *Weighing Reasons*, edited by Errol Lord and Barry Maguire, 27–55. Oxford: Oxford University Press, 2016.

Begon, Jessica. "Paternalism." *Analysis* 76, no. 3 (July 2016): 355–73.

Birks, David. "How Wrong Is Paternalism?" *Journal of Moral Philosophy* 15, no. 2 (April 2018): 136–63.

50 I am extremely grateful to Lorenzo Elijah and Jonathan Parry, who read multiple drafts and offered very useful feedback and discussion. I would also like to thank David Enoch, David Birks, Nikhil Krishnan, Jen Semler, Lewis Williams, Rose Macaulay, and the editors and reviewers of this journal for their helpful suggestions to improve this paper. A version of this paper was first presented at the Philosophy DPhil Seminar at the University of Oxford, where it benefitted from good feedback from the attendees. Finally, I owe thanks to the Institute for Ethics in AI at the University of Oxford for funding and supporting my research.

———. "Sex, Love, and Paternalism." *Ethical Theory and Moral Practice* 24, no. 1 (March 2021): 257–70.

Bullock, Emma C. "A Normatively Neutral Definition of Paternalism." *Philosophical Quarterly* 65, no. 258 (January 2015): 1–21.

Chang, Ruth. "Do We Have Normative Powers?" *Aristotelian Society Supplementary Volume* 94, no. 1 (July 2020): 275–300.

Cholbi, Michael. "Paternalism and Our Rational Powers." *Mind* 126, no. 501 (January 2017), 123–53.

Clarke, D. S., Jr. "Exclusionary Reasons." *Mind* 86, no. 342 (April 1977): 252–55.

Coons, Christian, and Michael Weber. "Introduction: Paternalism—Issues and Trends." In *Paternalism: Theory and Practice*, 1–24. Cambridge: Cambridge University Press, 2013.

Dancy, Jonathan. *Ethics without Principles*. Oxford: Oxford University Press, 2004.

Dworkin, Gerald. "Defining Paternalism." In *Paternalism: Theory and Practice*, edited by Christian Coons and Michael Weber, 25–38. Cambridge: Cambridge University Press, 2013.

———. "Paternalism." In *Morality and the Law*, edited by Richard Wasserstrom, 107–26. Belmont, CA: Wadsworth Publishing Company, 1971.

Enoch, David. "Hypothetical Consent and the Value(s) of Autonomy." *Ethics* 128, no. 1 (October 2017): 6–36.

———. "What's Wrong with Paternalism: Autonomy, Belief, and Action." *Proceedings of the Aristotelian Society* 116, no. 1 (April 2016): 21–48.

Feinberg, Joel. *Harm to Self*. Vol. 3 of *The Moral Limits of the Criminal Law*. Oxford: Oxford University Press, 1986.

Grill, Kalle. "Antipaternalism as a Filter on Reasons." In *New Perspectives on Paternalism and Healthcare*, edited by Thomas Schramme, 46–63. New York: Springer International Publishing, 2015.

———. "The Normative Core of Paternalism." *Res Publica* 13, no. 4 (December 2007): 441–58.

Grill, Kalle, and Jason Hanna, eds. *The Routledge Handbook of the Philosophy of Paternalism*. New York: Routledge, 2018.

Groll, Daniel. "Medical Paternalism—Part 1." *Philosophy Compass* 3, no. 9 (March 2014): 194–203.

———. "Paternalism and Rights." In Grill and Hanna, *The Routledge Handbook of the Philosophy of Paternalism*, 119–31.

———. "Paternalism, Respect, and the Will." *Ethics* 122, no. 4 (July 2012): 692–720.

Hanna, Jason. "Hard and Soft Paternalism." In Grill and Hanna, *The Routledge Handbook of the Philosophy of Paternalism*, 24–34.

———. *In Our Best Interest: A Defense of Paternalism*. Oxford: Oxford University Press, 2018.
Kleinig, John. *Paternalism*. Totowa, NJ: Rowman and Allanheld, 1983.
Malm, Heidi. "Feinberg's Anti-Paternalism and the Balancing Strategy." *Legal Theory* 11, no. 3 (September 2005): 193–212.
Mill, John Stuart. *On Liberty*. London: Penguin, 2010.
Moore, Michael S. "Authority, Law and Razian Reasons." *Southern California Law Review* 62, nos. 3–4 (1989): 827–96.
Owens, David. *Shaping the Normative Landscape*. Oxford: Oxford University Press, 2012.
Parfit, Derek. *Reasons and Persons*. Oxford: Oxford University Press, 1986.
Parry, Jonathan. "Defensive Harm, Consent, and Intervention." *Philosophy and Public Affairs* 45, no. 4 (September 2017): 356–96.
———. "What's Wrong with Paternalism?" Unpublished manuscript.
Quong, Jonathan. *Liberalism without Perfection*. Oxford: Oxford University Press, 2011.
Raz, Joseph. *The Morality of Freedom*. Oxford: Clarendon Press, 1986.
———. *Practical Reason and Norms*. Oxford: Oxford University Press, 1999.
Scanlon, T. M. "Reasons: A Puzzling Duality." In *Reasons and Value: Themes from the Moral Philosophy of Joseph Raz,* edited by R. Jay Wallace, Philip Pettit, Samuel Scheffler, and Michael Smith, 231–46. Oxford: Oxford University Press, 2004.
Shafer-Landau, Russ. "Liberalism and Paternalism." *Legal Theory* 11, no. 3 (September 2005): 169–91.
Shiffrin, Seana Valentine. "Paternalism, Unconscionability Doctrine, and Accommodation." *Philosophy and Public Affairs* 29, no. 3 (July 2000): 205–50.
Tsai, George. "Paternalism and Intimate Relationships." In Grill and Hanna, *The Routledge Handbook of the Philosophy of Paternalism*, 348–59.

Journal of Ethics and Social Philosophy
VOL. 26, NO. 3 · FEBRUARY 2024
https://doi.org/10.26556/jesp.v26i3.2719

MAXIM AND PRINCIPLE CONTRACTUALISM

Aaron Salomon

CONTRACTUALISM determines which actions I must perform by seeing whether they accord with principles for the general regulation of behavior that no one can reasonably reject.[1] Part of what makes contractualism such an attractive moral theory is its faithfulness to our concept of morality. It is part of our very idea of morality that it is *to be realized* by social institutions in that moral principles and rules are to be internalized by communities and regulate the activity of their members.[2] Contractualism's focus on evaluating principles for the general regulation of behavior allows it to vindicate our view that morality must be the kind of thing that can play a particular role in regulating the social order. Living together on the basis of principle contractualism's nonrejectable principles would be pretty good indeed. But this otherwise attractive feature of contractualism gives rise to the ideal world problem. Sometimes, contractualism recommends acting in accordance with a principle that would be great if it were generally accepted but a nightmare to follow in situations where it is not.

1 This paper's title is a play on Sheinman's "Act and Principle Contractualism." In that paper, Sheinman also argues that contractualists should no longer determine which actions I must perform by seeing whether they accord with certain principles for the general regulation of behavior. But there are two points of difference between my paper and his. First, Sheinman argues that contractualists ought to evaluate my acts *directly* for their rejectability or lack thereof in order to determine whether I am required to perform them. I, however, think that contractualists ought to assess any maxim that might be reflected in my actions for rejectability in order to determine whether I am required to perform them. Second, I argue that contractualists should drop their commitment to evaluating principles for the general regulation of behavior in order to solve the ideal world problem. In doing this, I follow Murphy, "Nonlegislative Justification." But Sheinman argues that contractualists should drop this commitment of theirs in order to allow their view to remain consistent with what he calls "foundational contractualism," the view that what matters ultimately in action is unrejectability. More on Sheinman's view later.

2 I owe my formulation of this plausible conceptual claim about morality to Walden, "Mores and Morals," 419. Walden traces this conceptual claim back to Hegel, Marx and Engels, and Nietzsche.

Suppose that if the principle "give to poverty relief if you are not impoverished yourself" were generally accepted, then global poverty would be alleviated. Suppose further that this principle is not generally accepted, and as a result, poorly funded charities do more harm than good with the money that is given to them. This is because those who run these poorly funded charities have responded, and will continue to respond, quite vindictively to the fact that it is not customary for the financially comfortable to give to poverty relief. Call this possible world "Vindictive World."[3]

Even from the perspective of Vindictive World, the principle "give to poverty relief if you are not impoverished yourself" is nonrejectable. That is because, by hypothesis, if such a principle were generally accepted, then global poverty would be alleviated. So, according to contractualism, those who are financially comfortable in Vindictive World are required to give to poverty relief. But this is the wrong result! Surely, a financially comfortable person in Vindictive Word should not give to poverty relief if doing so will do more harm than good (by contributing money to a malevolent organization). So, contractualism is not true in at least one possible world. If contractualism is true, then it is necessarily true (i.e., true in all possible worlds). So, it is not true. This is the *ideal world problem*.

In order to solve the ideal world problem while remaining faithful to our concept of morality, contractualists should no longer determine which actions I must perform by seeing whether they accord with certain principles for the general regulation of behavior. Instead, contractualists should determine whether it is right or wrong for me to perform an action by evaluating any *maxim* that might be reflected by my action. Often, when we act intentionally we have a maxim.[4] Maxims are reflected in our actions, and they are the principles according to which we see ourselves as acting. A maxim expresses a person's policy, or in cases where one has no settled policy, the principle underlying the particular intention or decision on which one acts.[5] From here on out, I will refer to contractualism in its classical form as "principle contractualism" and my amended version as "maxim contractualism." According to maxim contractualism, an agent's action is morally required under the circumstances just

3 This is a version of the "utility landmine" case in Podgorski, "Wouldn't It Be Nice?" The general form of Podgorski's case is this: some great good can be brought about if *x* percent or more of us do *A*. But if less than *x* percent of us do *A*, *A*-ing would be counterproductive, or it would in some way produce bad results.

4 I say "often when" instead of "anytime" to leave room for weak-willed actions that, though intentional, are paradigm examples of actions we do in spite of the policies we have adopted. For discussion of this feature of weak-willed actions, see Gressis, "Recent Work on Kantian Maxims 1," 223.

5 O'Neill, "A Simplified Account of Kant's Ethics."

in case any maxim that he might adopt that involves not performing that action under the circumstances is one that someone could reasonably reject.

Maxim contractualism does not require financially comfortable residents of Vindictive World to give to malevolent charities. If I, a well-off resident of Vindictive World, adopted, as a settled policy, the maxim of giving to poverty relief if I am not impoverished myself, then the money I donate will do more harm than good. And that is enough to make my maxim rejectable. This is the main idea. Refinements will follow.

Here is the plan. In section I, I present principle contractualism and highlight one of its central advantages—namely, its ability to "defend the moral moderate," as Rahul Kumar would put it, about beneficence, or charity.[6] In section II, I show how the very feature of principle contractualism that allows it to "defend the moral moderate" also makes it succumb to the ideal world problem. In section III, I present and reject one way that a contractualist might go in order to solve her ideal world problem while retaining the spirit of her view—namely, the adoption of act contractualism. Although act contractualists are right to drop the principle contractualist commitment to evaluating principles for the general regulation of behavior, their view fails. For, it cannot account for the fact that, sometimes, what would happen if I performed an action over time is relevant to whether I am permitted to perform that action right here, right now. Instead, as I argue in section IV, contractualists should determine whether it is right or wrong for me to perform an action by evaluating any maxim that might be reflected by my action. Section V compares maxim contractualism to a distinct version of contractualist moral reasoning, which has recently been defended by Liam Murphy in order to illustrate the importance of *ending*, rather than beginning, one's moral reasoning with the evaluation of general principles. In section VI, I anticipate some objections to maxim contractualism and respond to them. The resulting picture is that maxim contractualism is uniquely positioned to both solve the ideal world problem and vindicate the moral force of the question, "What if I did that over time?"

I

Principle contractualism is the view that an action is morally required just in case any principle for the general regulation of behavior that permitted people not to perform that action is one that someone could reasonably reject.[7] Principles for the general regulation of behavior say of *everyone* that they are

6 For Kumar's use of this phrase, see "Defending the Moral Moderate."

7 Scanlon, *What We Owe to Each Other*, 4.

permitted, required, or forbidden to perform certain actions. Consider, for example, Principle *F*—the principle, according to T. M. Scanlon, that explains promissory obligation:

> If (1) *A* voluntarily and intentionally leads *B* to expect that *A* will do *x* (unless *B* consents to *A*'s not doing *x*); (2) *A* knows that *B* wants to be assured of this; (3) *A* acts with the aim of providing this assurance, and has good reason to believe that he or she has done so; (4) *B* knows that *A* has the beliefs and intentions just described; (5) *A* intends for *B* to know this, and knows that *B* does know it; and (6) *B* knows that *A* has this knowledge and intent, then, in the absence of some special justification, *A* must do *x* unless *B* consents to *x*'s not being done.[8]

In front of Principle F, there are two implicit universal quantifiers that bind variables *A* and *B* and range over agents.

Someone can reasonably reject a principle for the general regulation of behavior just in case that principle is not acceptable, or justifiable, to every individual. A principle is not acceptable, or justifiable, to every individual just in case either

> (i) the reason that some individual has for objecting to the principle on the basis of its implications is stronger than the reasons that all other individuals have for wanting the kinds of normative powers, benefits, or protections secured by the principle, or
> (ii) there is some alternative principle that answers to a sufficient degree the reasons that the relevant individual has for favoring the original principle but whose implications do not justify objections to it from any individual that are as serious as those justified by the original principle's implications.[9]

What does it mean, however, to speak of a principle's "implications?" Whether or not one of these principles for the general regulation of behavior is one that someone could reasonably reject is determined by considering, from a variety of points of view, the effects of that principle's general acceptance. As Scanlon writes,

> When we think of those to whom justification is owed, we naturally think first of the specific individuals who are affected by specific actions. But when we are deciding whether a given principle is one that could reasonably be rejected, we must take a broader and more

8 Scanlon, *What We Owe to Each Other*, 304.

9 Scanlon, *What We Owe to Each Other*, 95.

> abstract perspective. This perspective is broader because, when we are considering the acceptability or rejectability of a principle, we must take into account not only the consequences of particular actions, but also the consequences of general performance or nonperformance of such actions and of the other implications (for both agents and others) of having agents be licensed and directed to think in the way that that principle requires. . . . [An] assessment of the rejectability of a principle must take into account the consequences of its *acceptance in general*, not merely in a particular case that we may be concerned with (emphasis mine).[10]

So, when we want to know whether a principle is one that someone could reasonably reject, we need to imagine what would happen if people generally governed their practical reasoning in terms of that principle. In other words, when we are interested in the rejectability of a principle, we need to turn our attention to the effects of its *internalization*.

Principle contractualism's focus on the effects of a principle's general acceptance provides it with the materials to vindicate a moderate position about when beneficence is required. Suppose I have an extra twenty dollars lying around. I can either spend that money at a movie theater or I can donate it to a local charity that is certain to feed someone with it who is down on their luck. Intuitively, so long as I am beneficent on occasion, I am permitted, in this instance, to spend my twenty bucks at the movies.[11] But how can this be accounted for if my enjoying a movie is less important than someone receiving help? Principle contractualism can get the right result in this case by pointing to the problems faced by those agents who govern their practical reasoning in terms of a principle that requires them to always do what is necessary to prevent another from incurring a significant loss, provided that they can do so at a cost to themselves that is less significant. If people reasoned about what to do in terms of such a principle, they would not have the kind of control over the course of their lives sufficient for making and executing plans. Nor would they have enough of the course of their lives dictated by the choices that they made.[12] Principle contractualism, precisely because it evaluates principles on

10 Scanlon, *What We Owe to Each Other*, 202–4.

11 In other words, although I may have what some have called an "imperfect duty" of beneficence, or charity, I am permitted in some instances to keep my extra money for myself. Of course, some philosophers would deny this. On their view, our obligations to the poor are much more demanding than commonsense morality would have them be. In this connection, see, for example, Kagan, *The Limits of Morality*, ch. 1.

12 This is how Kumar argues against a principle for the general regulation of behavior that requires one to always do what is necessary to prevent another from incurring a significant

the basis of the effects of their general acceptance, can avoid being as *demanding* as some of its consequentialist competitors.

II

The very feature of principle contractualism, however, that allows it to defend the moral moderate—namely, its focus on the effects of a principle's general acceptance—gives rise to extensional problems of its own. As we saw above, principle contractualism requires financially comfortable residents of Vindictive World to give to malevolent, charitable organizations *precisely because* it determines whether a well-off resident of Vindictive World should do so by looking at the implications of the principle "give to poverty relief if you are not impoverished yourself" being generally accepted. The implications would be good indeed.

Now that the machinery of principle contractualist moral reasoning is more clearly in view, let us rehearse this argument against principle contractualists: any principle that permitted someone who is financially comfortable not to give to poverty relief is one that someone could reasonably reject. From the perspective of Vindictive World, there is *really* strong reason not to want general acceptance for any principle that permitted everyone who is financially comfortable not to give to poverty relief. If any such principle is generally accepted, then poverty would *not* be alleviated. But there is no similarly strong reason to want any principle to be generally accepted that permitted someone who is financially comfortable not to give to poverty relief. Not wanting to have to give a little bit of money to poverty relief when one is financially comfortable

---

loss, provided that they can do so at a cost to themselves that is less significant. For his discussion, see "Defending the Moral Moderate," 296–303. Kumar also points out that principle contractualism is able to reject such a principle precisely because it determines whether we must do *A* by imagining what would happen if a principle that required us to do *A* had the status of *custom*. It is worth noting here, however, that some have dissented from the idea that principle contractualism can recognize the rejectability of the principle that requires one to always do what is necessary to prevent another from incurring a significant loss, provided that they can do so at a cost to themselves that is less significant. In this connection, see Ashford, "The Demandingness of Scanlon's Contractualism"; Hills, "Utilitarianism, Contractualism, and Demandingness." But these arguments target the relative importance of the interest in control that purportedly grounds the rejectability of the principle in question, not whether looking at worlds where a principle is generally accepted allows principle contractualists to identify such an interest. These arguments, in other words, do *not* target the idea that principle contractualism's focus on the effects of the general acceptance of principles puts it in a better position to "defend the moral moderate" than theories that focus only on the effects of particular actions. As such, they are offstage dialectically for me.

pales in comparison to the kind of lives people would live were global poverty to be alleviated. As discussed earlier, this is a most unintuitive result. That is because a financially comfortable person in Vindictive World should not give to poverty relief if doing so will do more harm than good.

Now, it may be wondered why the following principle is rejectable: give to poverty relief if you are not impoverished yourself, *unless it is not customary to give to poverty relief, in which case do not*. This principle includes the circumstances in which the simpler principle—"give to poverty relief if you are not impoverished yourself"—is not generally accepted *in its own formulation*. For ease of exposition, let us call this principle the "complicated poverty principle," and let us call "give to poverty relief if you are not impoverished yourself" the "simple poverty principle." If the complicated poverty principle were nonrejectable, then principle contractualists would not face the ideal world problem as I have characterized it. For, if the complicated poverty principle were nonrejectable, then not only would a financially comfortable resident of Vindictive World not be required to give to poverty relief, they would be required *not* to do so. But is the complicated poverty principle nonrejectable?

No. Who stands to *gain* the most from the complicated poverty principle governing charitable giving in Vindictive World? The affluent do, since if they govern their practical reasoning in terms of the complicated poverty principle, then they will not be required to do more harm than good with their money. Who, moreover, stands to *lose* the most from the complicated poverty principle governing charitable giving in Vindictive World? Those who are impoverished do. They could easily say: the complicated poverty principle is worse than the simple poverty principle because, if the simpler principle were generally accepted, then global poverty would be alleviated! Surely, not being impoverished is more morally important than efficiently using one's extra money. This makes the complicated poverty principle rejectable.

So, principle contractualism cannot get the right result in Vindictive World by including circumstances in which the principles are not generally accepted *in the statement of a principle*.[13] How might a contractualist alter her view in

13 Parfit, too, rejects this principle contractualist move, but he does so on very different grounds. His grounds for rejecting this move are different in part because his target is *not* a principle contractualism that evaluates principles for the general regulation of behavior in terms of the effects of their general acceptance. Instead, his target is a principle contractualism—namely "Kantian contractualism"—that evaluates principles in terms of the effects of their *universal compliance*. For his discussion, see *On What Matters*, 1:312–20.

Neither, in order to get the correct result about Vindictive World, can principle contractualists appeal to an alternative principle, such as "do not harm others" or "avoid disaster at all costs," that they may also take to be operative in Vindictive World and outweigh the poverty principle. And this is for similar reasons: it is false that "avoid disaster at all

order to avoid the ideal world problem? And how can she do so while retaining her ability to defend a moderate view about when beneficence is required?

III

Abelard Podgorksi has helpfully observed that the ideal world objection "faces any view which determines what we individuals ought to do in this world by evaluating worlds that differ from the actual world in more than what is up to us."[14] Principle contractualism certainly evaluates possible worlds that differ from the actual world in more than what is up to us. For, it is not up to me whether a candidate moral principle is generally accepted. So, perhaps we should alter contractualism such that it evaluates worlds that differ from the actual world *in only what is up to us*.

One way of doing this would be to make contractualism's primary evaluative focal points *actions* rather than principles for the general regulation of behavior.[15] To do this would be to adopt act contractualism, according to which an agent's action is morally required just in case someone could reasonably reject that agent's not performing that action.[16] Someone can reasonably reject another's not performing an action just in case that other person's omission is not acceptable, or justifiable, to every individual. An agent's omission is not acceptable, or justifiable, to every individual just in case the reason that some individual has for objecting to the omission on the basis of its implications is stronger than the reason that the agent (or a third party) has for wanting the benefits she (or a third party) would get from the omission.[17] Act contractual-

costs" applies to our protagonist in Vindictive World. That is because, from the perspective of Vindictive World, "avoid disaster at all costs, unless you are causing disaster by giving to poverty relief" is rejectable. Imagine a world that is just like Vindictive World except that people start governing their practical reasoning in terms of "avoid disaster at all costs, unless you are causing disaster by giving to poverty relief." In such a world, the charities would solve poverty relief. Why? Well, because, in that world, generally, people accept a principle that involves giving to those malevolent charities. And malevolent charities respond by eradicating poverty relief. For an argument with a similar conclusion concerning which moves are open to the rule consequentialist to solve its ideal world problem, see Podgorski, "Wouldn't It Be Nice?," 286.

14 Podgorski, "Wouldn't It Be Nice?," 279. Of course, such a diagnosis of what makes a normative ethical theory face the ideal world problem is *defeasible*, in the sense that it is very plausible but may prove to be too quick in light of forthcoming principle contractualist (or rule consequentialist) attempts to get the right result about charity in Vindictive World.

15 I borrow the phrase "evaluative focal point" and its cognates from Kagan, "Evaluative Focal Points."

16 Sheinman, "Act and Principle Contractualism," 295.

17 Sheinman, "Act and Principle Contractualism," 296.

ism, then, evaluates worlds that differ from the actual world in only what is up to us since it only evaluates worlds in which we do not perform some action.

This feature of act contractualism allows it to get the right result in Vindictive World. Act contractualism tells me, a financially comfortable inhabitant of Vindictive World, *not* to give to poverty relief for someone could reasonably reject my giving to poverty relief. From the perspective of Vindictive World, there is *really* strong reason to not want me to give to poverty relief. If I did give to poverty relief, those who run the poorly funded charities would use the money that I gave them to do more harm than good. But I do not have similarly strong reason to want to give to poverty relief. If the whole point of giving to charity was to do good, then what is the point of giving to charity when doing so will do more harm than good?

Act contractualism, of course, has a structure similar to a much more familiar view called "act consequentialism," according to which an agent's action is morally required just in case it would result in more well-being overall than any of the other actions available to the agent.[18] However, there are a few key differences between these two normative ethical theories. For one, act contractualism is better placed than act consequentialism is to account for our intuition about Scanlon's Transmitter Room case. Suppose that Jones has suffered an accident in a TV broadcasting station and is receiving extremely painful electrical shocks. If we turn off the power to save him, billions of viewers will miss the last half hour of the World Cup final. Intuitively, it would be wrong not to save Jones from his agony, regardless of how many people are watching the game. The benefit of watching a soccer match is trivial compared to the agony of suffering strong electrical shocks. No matter how large the sum of these benefits, it would seem wrong to keep the power on. Act consequentialism has trouble vindicating this intuition since it allows for the interpersonal aggregation of well-being. It seems like the act consequentialist is forced to agree that, at some point, the combined benefits to the viewers must become large enough to morally outweigh Jones's agony.[19]

On the other hand, act contractualism is able to get the right result in this case since it retains two of the restrictions that principle contractualism places on the reasons that can be pressed for and against candidate moral principles in contractualist moral reasoning. First, act contractualism retains the *impersonalist restriction*, according to which one cannot appeal to claims about the impersonal goodness or badness of outcomes when one is rejecting or favoring

18 Kagan, "Evaluative Focal Points," 134.

19 Scanlon, *What We Owe to Each Other*, 235.

some candidate moral principle.[20] For the act contractualist, of course, this restriction looks a bit different since its primary evaluative focal points are *actions* rather than principles for the general regulation of behavior. The act contractualist impersonalist restriction, then, says that one cannot appeal to claims about the impersonal goodness or badness of outcomes when one is rejecting or favoring some action. So, the sum of benefits that those who are watching the World Cup final will enjoy if we keep the power on will not even enter into act contractualist moral reasoning. To point out such a sum when arguing that it is nonrejectable or unrejectable to keep the power on would be to appeal to a claim about the impersonal goodness of an outcome.

Second, act contractualism retains the *individualist restriction*, according to which, when one is rejecting or favoring some candidate moral principle, one must only appeal to that principle's implications for ourselves and for other particular people.[21] For the same reasons noted above, the act contractualist version of this restriction will look a bit different: when one is rejecting or favoring some action, one must only appeal to that action's implications for ourselves and for other particular people. So, the act contractualist will compare the reasons that Jones has to avoid terrible suffering with the reasons a *particular* World Cup final watcher will have to want to enjoy the game, finding the reasons Jones has to be much stronger. Act contractualism, roughly speaking, is act consequentialism *plus the individualist and impersonalist restrictions*.

Despite its focus on acts rather than principles, act contractualists may even be able to vindicate our concept of morality as a social institution by generating *rules of thumb*. It is not always easy to correctly determine whether or not the performance of a particular action in a particular circumstance is unrejectable. But we also need to deliberate about what to do in terms of facts about whether an action is unrejectable. After all, moral considerations carry great weight in the practical deliberations of a virtuous agent. And act contractualism, in its capacity as a moral theory, seeks to establish these very sorts of considerations. The solution is to establish rules such as "one may do *A* in *C*" if, in most circumstances, performing some action is unrejectable.[22] Adopting such rules of thumb allows agents to be guided by the realization of unrejectability in their practical reasoning without needing to assess in every circumstance which one of the actions available to them is unrejectable. If these rules of thumb become internalized by communities and regulate the activity of their members, then

20 Parfit, *On What Matters*, 2:214; Scanlon, *What We Owe to Each Other*, 222.

21 Parfit, *On What Matters*, 2:193; Scanlon, *What We Owe to Each Other*, 229.

22 For analogous comments concerning act consequentialism, see Rawls, "Two Concepts of Rules," 18–29.

those communities will be marked by a particularly high level of unrejectability. This will lead to a sort of social harmony in those places since their members will mostly live together in ways that they can justify to each other.[23]

Act contractualism, however, is too demanding. Recall a case from section 1: suppose I have an extra twenty dollars lying around. I can either spend that money at a movie theater, or I can donate it to a local charity that is certain to feed someone with it who is down on their luck. According to act contractualism, it is not permissible for me to go to the movies since the down-on-their-luck person has stronger reason to want me to give the money to the charity than I have reason to want to go to the movies. The benefit of watching a movie and eating some popcorn is trivial compared to the suffering involved in starvation. Intuitively, however, (as mentioned in section 1) it *does* seem permissible for me to go to the movies (since it seems permissible for me to either go to the movies or give to the charity). So, act contractualism cannot account for what is common sense—that sometimes I am permitted not to bring about unrejectability. What ground act contractualism seems to gain over principle contractualism by solving the ideal world problem, it loses by being unable to defend a moderate position about the conditions under which charity is required.

What is more, this problem with act contractualism goes *deeper* than its demandingness. What is preventing act contractualism from being able to account for the permissibility of my spending my twenty dollars at the movies? It is the fact that act contractualism is built not to recognize that the cumulative intrapersonal burdens of my acting in a certain way over time can make acting in that way merely permissible rather than required. Act contractualism only considers my reason to want to act in some way on a particular occasion. But what seems to make going to the movies permissible is that, if I had to always give any extra money I had to charity, then I would not have sufficient control over the course of my life that I am able to make and execute plans and to some extent have the course of my life dictated by the choices I make.

So, it seems like act contractualism is too demanding because it is unable to account for the fact that sometimes what would happen if I performed an action over time is relevant to whether I am permitted to perform that action right here, right now. In other words, it seems like there is a certain *kind* of moral objection that one might make to an action which act contractualist moral reasoning cannot capture. And we can confirm this by looking at more examples.

23 For a familiar defense of the connection between unrejectability and justifiability to another, see Scanlon, *What We Owe to Each Other*.

The demandingness objection to act contractualism concerned a case with the following structure: *X*'s *A*-ing would be morally permissible; the reasons *X* has against *X*'s *A*-ing on multiple occasions over time are stronger than *Y*'s reasons to want *X* to do *A* over time, but the reasons *Y* has to want *X* to do *A* on a particular occasion are stronger than the reasons *X* has against *X*'s doing *A* on that occasion. But there are also cases with this structure: *X*'s *A*-ing would be morally wrong; the reasons a distinct agent *Y* has against *X*'s *A*-ing on multiple occasions over time are stronger than *X*'s reasons to want to *A* over time, but the reasons *X* has to want to do *A* on a particular occasion are stronger than the reasons *Y* has against *X*'s doing *A* on that occasion.

Here is one such case: suppose I stand up my friend for a coffee date in order to go to a talk on normative ethics. This is the only time I have stood him up, so it does not cause him any psychological harm. Nor does my standing him up inconvenience him. He was planning on working at the coffee shop we agreed to meet at after we met, and if we had not had a coffee date scheduled, he would have just come to the coffee shop earlier to work. Nonetheless, it seems that, by standing up my friend for a coffee date, I have done wrong. But act contractualism does not seem to be able to capture this intuition. This is because I have stronger reason to want to stand up my friend than he has reason to want me not to. I would benefit from going to a talk in my field, and my friend would not really be harmed at all by my standing him up.

An act contractualist may reply that, in agreeing to go on a coffee date with my friend, I have promised to meet him at coffee shop *X* on occasion *Y*. So, he may continue, even if I do not cause him any psychological harm when I stand him up on *Y*, I do cross on *Y* whatever interest of his it is that grounds promissory obligation in act contractualist moral reasoning—perhaps his interest in others doing what they assured him they would do.[24]

But it is not plausible that the wrong I have committed by standing up my friend is the wrong of breaking a promise. Here is a datum about promissory obligation: one cannot fulfill one's duty to keep one's promise to do *A* by warning the promisee before she has undertaken an action based on one's promise that one will not, after all, do *A*. To see that this is true, suppose that I promise to drive you to work if you mow my lawn, and you accept. Then, a day later (but before either of us has begun doing what we promised the other to do), I change my mind and try to back out by warning you that I will not drive you to work even if you do, in fact, mow my lawn. Intuitively, by warning you that I will not drive you to work even if you hold up your end of the bargain, I have

24 Scanlon, *What We Owe to Each Other*, 303–4.

not made it such that I am no longer obligated to drive you to work if you mow my lawn.[25]

Thus, if my wrong were the same kind of wrong as the wrong of promise-breaking, then I could not have extinguished my duty to meet up with my friend by giving him a timely warning that I will not be able to make it. Intuitively, though, this is false. I do not wrong my friend by not showing up to the coffee shop when I warn him that I will not before he has undertaken any action based on our arrangement (such as, say, turning down another friend's suggestion to meet up).[26]

What is preventing act contractualism from being able to account for the wrong of my standing up my friend (assuming, that is, that it cannot—there may be some other interest I have not canvassed that grounds the nonrejectability of my harmlessly standing up my friend)? It is the fact that act contractualism is built not to recognize that the cumulative *inter*personal burdens of my acting in a certain way over time can make acting in that way wrong. Act contractualism only considers objections to my acting on a particular occasion. But the decisive objection to my standing up my friend to go to a talk seems to be that, if I were to stand him up on many occasions over time, I might cause him great psychological harm. He may come to feel like I have no respect for him, like I do not value him at all.

The fact that there are cases with these structures (i.e., the movie and coffee shop cases) suggests that what would happen if I performed an action over time can be relevant to whether I am permitted to perform that action right here, right now. How might we capture this in contractualist moral reasoning without courting the ideal world objection? We need a version of contractualist moral reasoning, then, which both (i) acknowledges that what would happen if I performed an action over time can be relevant to whether I am permitted to perform that action right here, right now, and (ii) evaluates worlds that differ from the actual world in only what is up to us.

IV

In order to generate such a version of contractualist moral reasoning, we should alter contractualism such that its primary evaluative focal point is a *maxim*. The resulting view is maxim contractualism, according to which an agent's action is morally required under the circumstances just in case any maxim that he might

25 I borrow this example from Scanlon, *What We Owe to Each Other*, 301.

26 This problem for the act contractualist response that ties the wrong of my standing up my friend to the wrong of breaking a promise also faces the act contractualist response that ties the wrong of my standing up my friend to the *mere* disappointment of his expectations.

adopt that involves not performing that action under the circumstances is one that someone could reasonably reject.[27] Whether or not one of these maxims is one that someone could reasonably reject is determined by considering the implications of the agent being guided in his practical reasoning over time by that maxim from a variety of points of view.[28] We will see imminently why maxim contractualism does not invite the ideal world objection. But now we can see why it will be able to acknowledge that what would happen if I performed an action over time can be relevant to whether I am permitted to perform that action right here, right now. For according to maxim contractualism, the fundamental moral question is: "What if I did that over time?"

How exactly does maxim contractualist moral reasoning work? Someone can reasonably reject a person's maxim just in case that maxim is not acceptable, or justifiable, to every individual. His maxim is not acceptable, or justifiable, to every individual just in case either

> (i) the reason that some individual has for objecting to the maxim on the basis of the implications of the agent being guided in his practical reasoning over time by that maxim is stronger than the reasons that the agent has for wanting the kinds of benefits he gets from being guided in his practical reasoning over time by that maxim, or

27 The fact that maxim contractualism's objects of moral assessment are all the maxims that my action *might* be in accordance with allows it to be squared with a central commitment of Scanlon's—the *Irrelevance-of-Intention-to-Permissibility Thesis* (Scanlon, *Moral Dimensions*). According to this thesis, it is irrelevant to the question of whether *X* may do ϕ what intention *X* would ϕ with if he or she did it (Thomson, "Self-Defense," 294). Maxim contractualism is consistent with the Irrelevance-of-Intention-to-Permissibility Thesis because its moral reasoning does not involve evaluating the *actual* intention with which an agent performed some action. Rather, the moral reasoning distinctive of maxim contractualism involves evaluating any intention with which the agent *might* have performed the action in question. This provides one source of contrast between maxim contractualism and another normative ethical theory that assesses maxims: universal law Kantianism. For, universal law Kantianism asks us, when evaluating an agent's action, whether the actual intention that the action is in accordance with could be willed by the agent to be a universal law.

28 In other words, whether or not one of these maxims is one that someone could reasonably reject is determined by considering the implications of him *in fact* adopting that maxim as a settled policy from a variety of points of view. I emphasize "in fact" because one's adopting a maxim does not necessarily involve sticking with it on multiple occasions. I may make a lying promise on the basis of the following maxim: "When I believe myself to be in need of money I shall borrow money and promise to repay it, even though I know that this will never happen." (Kant, *Groundwork of the Metaphysics of Morals*, 4:422.)

Then, after reading Kant's *Groundwork*, I may come to agree with him that no one could will that such a maxim become a universal law and no longer govern my practical reasoning in terms of that maxim.

(ii) there is some alternative maxim the agent's adoption of which over time answers to a sufficient degree the reasons that the agent has for favoring his adoption of the original maxim over time but whose implications do not justify objections to it from any individual that are as serious as those justified by the original maxim's implications.

As should be evident, maxim contractualist moral reasoning is roughly the same as principle contractualist moral reasoning. The main difference is that maxim contractualism evaluates worlds that differ from the actual world in only what is up to the agent since it only evaluates worlds in which the agent regulates her behavior over time in terms of a particular maxim that she is able to adopt.[29] And this makes all the difference for solving the ideal world problem.[30]

Like act contractualism, maxim contractualism allows us to get the right result in Vindictive World. Remember what is going on in Vindictive World:

> If the principle "give to poverty relief if you are not impoverished yourself" were generally accepted, then global poverty would be alleviated. But this principle is not generally accepted. And as a result, poorly funded charities do more harm than good with the money that is given to them. This is because those who run these poorly funded charities have and will continue to respond quite vindictively to the fact that it is not customary for the financially comfortable to give to poverty relief!

29 This brings out the importance of refraining from reading "might" as indicating metaphysical possibility in my formulation of maxim contractualism: an agent's action is morally required under the circumstances just in case any maxim that he might adopt that involves not performing that action under the circumstances is one that someone could reasonably reject. This is because if "any maxim he might adopt" were read as any maxim that it is metaphysically possible for him to adopt, then maxim contractualism would *not* evaluate worlds that differ from the actual world in only what is up to the agent. For there may be some maxims that an agent is not *able* to adopt as a settled policy even though it is metaphysically possible for him to adopt them as a settled policy. But maxim contractualism must retain its commitment to only evaluating worlds that differ from the actual world in only what is up to the agent. As we are about to see, that is the commitment that allows maxim contractualism to avoid the ideal world objection.

30 Another version of contractualism that might go by the name of "maxim contractualism" says that an agent's adoption of a maxim is morally required just in case any principle that permitted her not to adopt that maxim is one that someone could reasonably reject. Contractualists should not adopt this kind of maxim contractualism, however, since it will not help them solve their ideal world problem. That is because, like principle contractualism, the version of maxim contractualism under discussion in this footnote determines what I ought to do (or at least which maxims I ought to adopt) by evaluating worlds that differ from the actual world in more ways than are up to me.

Maxim contractualism does not require me, a financially comfortable denizen of Vindictive World, to give to poverty relief. Roughly speaking, maxims take the following form: in circumstances *C*, I shall do *A*.[31] On my usage of "maxim," then, maxims are something like principles for the regulation of *my* behavior. So, one maxim that involves doing what the principle "give to poverty relief if you are not impoverished yourself" requires in the circumstances in which that principle requires it is: "If I am financially comfortable, then I shall give to poverty relief." If I adopt this maxim as a policy over time, then I will do more harm than good with my money. And I do not have any reason to give to poverty relief other than to do net good with my money. So, one of the maxims that involves doing what the principle "give to poverty relief if you are not impoverished yourself" requires is rejectable. So, I, a financially comfortable resident of Vindictive World, am not required to give my money to malevolent organizations.

Moreover, also like act contractualism, maxim contractualism tells me, a financially comfortable denizen of Vindictive World, *not* to give to poverty relief. For, any maxim that I may adopt over time that involves giving to poverty relief is one that someone could reasonably reject. From the perspective of Vindictive World, there is *really* strong reason to not want me to adopt any maxim over time that involves me giving to poverty relief. If I adopted such a maxim over time, those who run the poorly funded charities would gain a

31 I say "roughly speaking" because, strictly speaking, maxims take the following form: when in circumstances *C*, I shall perform act *A in order to achieve end E* (Korsgaard, "Acting for a Reason," 219). This more precise description of the form that maxims take allows us to see why it is plausible to think that, often when we act intentionally, we have a maxim. It is plausible that most intentional action is *purposive*, in the sense that intentional action is not mere behavior. Intentional action is not just some string of observable events in the external world. Rather, it involves an agent willing that some end or purpose be achieved (Korsgaard, "Morality as Freedom," 162–67). So, often when we act intentionally, it is plausible that behind our action lies a maxim specifying the end that we are trying to realize in so acting. (Again, I say "often when we act intentionally" rather than "anytime we act intentionally" to leave room for weak-willed actions.) Nonetheless, I will stick with my looser characterization of maxims since the more precise form that a maxim takes is not relevant to my arguments in this essay.

As mentioned in the paragraph above, my rough characterization of the form that maxims take follows Korsgaard's influential reading of what Kant has in mind when he uses the word "maxim." But I do not intend my characterization to be a *reading* of Kant. Perhaps by "maxim" Kant had in mind something else. Perhaps he had in mind a kind of principle so broad that it cannot accommodate the specification of circumstances. That would be fine. All I need, for my purposes, is that the primary evaluative focal point of contractualist moral reasoning be a principle that regulates an individual's activity over time *and only that individual's activity*. What Korsgaard calls a "maxim" is precisely that. I could have called it a "principle for the regulation of *my* behavior" instead.

decent amount of money from me, with which they would do quite a bit more harm than good. But I do not have similarly strong reason to want to adopt a maxim over time that involves me giving to poverty relief. If the whole point of regularly giving to charity was to regularly do good, then what is the point of regularly giving to charity when doing so often will do much more harm than good?

Unlike act contractualism, however, maxim contractualism is able to capture the fact that, sometimes, what would happen if I performed an action over time is relevant to whether I am permitted to perform that action right here, right now. First, maxim contractualism can recognize the fact that I am permitted to spend my twenty dollars at the movies. Maxim contractualism determines whether I am permitted to spend my twenty dollars at the movies (in part) by assessing the significance of my adoption of a principle bearing on whether I do so. Suppose I adopted the following principle of mutual aid as my maxim:

> *Mutual Aid*: I will always do what is necessary to prevent another from incurring a significant loss, provided that I can do so at a cost to myself that is less significant.[32]

The intrapersonal burdens of governing my practical reasoning in terms of this maxim are much weightier than the burden of not being able to go to the movies on a single occasion. I have strong reason to want to have sufficient control over the course of my life that I am able to make and execute plans and, to some extent, have the course of my life dictated by the choices I make. My governing my deliberations in terms of Mutual Aid would prevent me from having the relevant kind of control. Moreover, it is plausible that such an intrapersonal burden is weighty enough to make Mutual Aid rejectable in comparison to a maxim that involves me only giving my fair share.

As I argued in section 1, this stretch of moral reasoning, with some tweaks, could easily be adopted by a principle contractualist. Like maxim contractualism, principle contractualism could determine whether I am permitted to spend my twenty dollars at the movies (in part) by assessing the significance of my adoption of a principle bearing on whether I do so. What distinguishes principle contractualism from maxim contractualism in this regard is only that the kind of principle that principle contractualism imagines me adopting is a principle for the general regulation of behavior rather than a maxim. It is a virtue of maxim contractualism that it is able to retain principle contractualism's ability

32 This is the maxim version of a principle (for the general regulation of behavior) of mutual aid that Kumar considers and rejects, with principle contractualist moral reasoning, in "Defending the Moral Moderate," 296–303.

to "defend the moral moderate" without succumbing to principle contractualism's ideal world problem. And, as we saw above, act contractualism is only able to do the latter of those things.

Second, maxim contractualism is able to recognize the fact that it is wrong for me to stand up my friend (even harmlessly). Suppose I adopted the following maxim:

> *Stand Up*: I will stand people up when it is convenient for me to do so, provided that my standing them up does not harm them.

The *inter*personal burdens that my friend would experience as a result of me governing my practical reasoning over time in terms of Stand Up are much weightier than the burden my friend would experience as a result of being stood up on the occasion in question. If I were to stand him up over time, he would eventually feel like I did not care about him, like he mattered not at all to me. And my friend has a very weighty interest in avoiding being thought of in such a way by his friends. Moreover, it is plausible that such an interpersonal burden is weighty enough to make Stand Up rejectable in comparison to a maxim that involves me never standing up others (even harmlessly).

However, we, of course, want our moral theory to be able to tell us not only that *I* am permitted or required (not) to perform some action (such as spending my twenty dollars at the movies or standing up my friend), but also whether *everyone* is. If it is part of our concept of morality that moral principles and rules are to be internalized by communities and regulate the activity of their members, then this desire makes sense. If our moral theory did not vindicate principles that are bound at a high enough level of generality, *we* could not live together on the basis of them, though maybe we could all hold someone (or a few people) accountable for certain actions on the basis of them. So, we want our moral theory to be able to make true sentences like, "Promises must be kept." It is natural to think that the way to do this, for the maxim contractualist, is to take each person and ask whether there is a nonrejectable maxim they might adopt over time, according to which they shall keep their promises under the relevant circumstances. And that may seem very unattractive. It would take a great deal of time indeed to establish this!

But maxim contractualists need not do this. If Principle *F* were my maxim, it would say:

> If (1) I voluntarily and intentionally lead *B* to expect that I will do *x* (unless *B* consents to my not doing *x*); (2) I know that *B* wants to be assured of this; (3) I act with the aim of providing this assurance, and have good reason to believe that I have done so; (4) *B* knows that I have

> the beliefs and intentions just described; (5) I intend for *B* to know this, and know that *B* does know it; and (6) *B* knows that I have this knowledge and intent, then, in the absence of some special justification, I shall do *x* unless *B* consents to *x*'s not being done.

If such a maxim were nonrejectable, then we would know that the following moral principle (for the regulation of *my* behavior) would be true:

> If (1) I voluntarily and intentionally lead *B* to expect that I will do *x* (unless *B* consents to my not doing *x*); (2) I know that *B* wants to be assured of this; (3) I act with the aim of providing this assurance, and have good reason to believe that I have done so; (4) *B* knows that I have the beliefs and intentions just described; (5) I intend for *B* to know this, and know that *B* does know it; and (6) *B* knows that I have this knowledge and intent, then, in the absence of some special justification, I must do *x* unless *B* consents to *x*'s not being done.

Now, a *nonsubstantive universalizability thesis* is a universalizability thesis that does not entail alone, or together with other nonmoral premises, any moral conclusions of the sort that something (some action, person, state of affairs) has a certain moral property.[33] Here is a very plausible example of one:

> If an action is right (or wrong) for one agent in a certain circumstance, then it is right (or wrong) for any similar agent in similar circumstances.[34]

This nonsubstantive universalizability thesis gives expression to the more general thought that "moral properties of things (persons, actions, states of affairs, situations) are essentially independent of their purely 'individual' or 'numerical' aspects."[35]

It follows from the truth of our nonsubstantive universalizability thesis and the truth of Principle *F* (understood as a principle for the regulation of my behavior) that Principle *F* (formulated with the generality that principle contractualists formulate it) is true. For, the relevant circumstances are those picked out in the antecedent of Principle *F* (understood as a principle for the regulation of my behavior): being such that you voluntarily and intentionally led another to expect that you will do *x* (unless that other consents to your not doing *x*), etc. Anyone who finds themselves in these circumstances must do what they assured another they would do. Maxim contractualism, then, can

33 Potter and Timmons, "Introduction," xii.

34 Potter and Timmons, "Introduction," xv.

35 Rabinowicz, *Universalizability*, 11.

vindicate our concept of morality as a social institution just as well as principle contractualism can.

V

We can see the importance of not attempting to universalize to general principles until the *end* of contractualist moral reasoning by reflecting on Liam Murphy's recent attempt to save contractualism from its ideal world problem. In "Nonlegislative Justification," Murphy defends a version of contractualism that is like principle contractualism in that it determines which actions I must perform by seeing whether they accord with principles that no one can reasonably reject. Where Murphy's view differs from principle contractualism is in the *kind* of principles that are to be assessed for reasonable rejection. As discussed, principle contractualists take principles for the general regulation of behavior to be the sorts of principles assessed for reasonable rejection during moral reasoning. According to Murphy, however, the kind of principles that are relevant are "general principles for cases like this in circumstances like these."[36] Take, for example, Murphy's Principle *R*, which he takes to be a paradigm example of a general principle for cases like this in circumstances like these:

> If one person invites another to rely on their stated intentions, and the other person does rely, then the first person must do what they can to prevent that reliance from coming at a loss.[37]

Murphy claims that *R* is a general principle for cases like this in circumstances like these rather than a principle for the general regulation of behavior because, unlike principles for the general regulation of behavior, *R* is not to be assessed for reasonable rejectability by imagining what would happen if it were generally accepted. As a general principle for cases like this in circumstances like these, *R* is to be assessed for (as we will see at least provisional) reasonable rejectability as follows: suppose I am considering whether I am required to prevent your invited reliance on me from coming at a cost to you. You might propose *R* and note the interest you have in my being required by *R* to do what will prevent your relying on me from being costly. Then, I might note my interest in not being required by *R* to prevent your relying on me from being costly. Then, we might come to see that your interest is stronger and so that *R* is (provisionally)

36 Murphy, "Nonlegislative Justification," 252.

37 Murphy, "Nonlegislative Justification," 255.

nonrejectable—making it (provisionally) the case that I am required to do what will prevent your relying on me from being costly.[38]

So, Murphy thinks that his view differs from principle contractualism in that whether a general principle is (provisionally) nonrejectable depends on the outcome of a conversation between two people in which they ask for and give each other reasons, rather than on what would happen if that principle were generally accepted. But, as Murphy correctly notes, once we decide that, e.g., *R* is nonrejectable, we have set a *precedent*. We have made a decision about how I ought to act whenever I have invited others to rely on my doing something. Murphy writes: "It is therefore appropriate, when offering reasons for and against a principle, to consider the possible cumulative 'intrapersonal' burdens it would entail."[39] (This is why the outcome of a conversation between two people in which they ask for and give each other reasons can only be that a general principle is *provisionally* nonrejectable.)

Murphy's claim seems right to me. If we are to set a precedent for me, we had better consider the effects of my deferring to that precedent going forward. However, when we decide that *R* is nonrejectable, we have not just set a precedent for me. We have set a precedent for *everyone*. *R* is a general principle for cases like this in circumstances like these, so *R* is a general principle that binds everyone. So, it is also appropriate, when offering reasons for and against *R*, to consider the possible cumulative intrapersonal burdens everyone's adoption of *R* would entail. But this would just be to consider the effects of the general acceptance of *R* while determining *R*'s rejectability, making Murphy's view determine what I ought to do (in part) by evaluating worlds that differ from the actual world in more than what is up to me and, thus, vulnerable to the ideal world objection.

Consider, by way of illustration, one of Gideon Rosen's examples: imagine a possible world—call it "Gremlin World"—that is just like ours except that it "contains a thing—a demon, or perhaps an inanimate device or natural force—that will wreak havoc if we attain moral unanimity. This gremlin is sensitive to our moral beliefs and abhors consensus. If we were to agree about some moral principle, it would cause universal misery, or destroy the world."[40] Now, it seems like, even in Gremlin World, when I have invited another to rely on my stated intentions, and the other person does rely, then I must do what I can to prevent that reliance from coming at a loss. As Rosen writes (though about a different sort of action):

38 Murphy, "Nonlegislative Justification," 255.

39 Murphy, "Nonlegislative Justification," 254.

40 Rosen, "Might Kantian Contractualism Be the Supreme Principle of Morality?," 84.

> One way to bring this out is to note that our world may, for all we know, contain a force that would wreak havoc if moral unanimity were attained. And yet the existence or non-existence of such a force is totally irrelevant to the moral assessment of ordinary human action in our world. An ordinary act of kindness is still right, even if somewhere in some cave some malignant thing is poised to spoil the universe if moral consensus is achieved.[41]

But we cannot vindicate this judgment using Murphy's preferred contractualist moral reasoning. *R* binds everyone, so Murphy thinks we need to imagine what would happen were everyone in Gremlin World to reason about what to do in terms of *R*. But if everyone in Gremlin World does *that*, the demon will destroy the world, making *R* rejectable.[42] It is precisely because Murphy's view determines what I ought to do (in part) by evaluating worlds that differ from the actual world in more than what is up to me that he cannot vindicate our moral judgment about Gremlin World. And it is precisely because he thinks contractualist moral reasoning ought to begin by assessing general principles (even ones not for the general regulation of behavior) that his view determines what I ought to do (in part) by evaluating worlds that differ from the actual world in more than what is up to me.

Maxim contractualism, on the other hand, can vindicate our judgment that, even in Gremlin World, it is wrong to not prevent your relying on me from being costly. My adoption of a maxim over time that involves preventing your relying on me from being costly would not result in the destruction of the world because the Gremlin only wreaks havoc when we attain moral unanimity. My practical consistency does not amount to our unanimity. Once again, the fact that maxim contractualism holds fixed everything that is outside my control when determining what my obligations are makes all the difference. And maxim contractualism is able to do this because it does not commence moral reasoning by considering general principles of any kind. The source of

41 Rosen, "Might Kantian Contractualism Be the Supreme Principle of Morality?," 84.

42 One might wonder why Murphy cannot reply that, from the perspective of Gremlin World, *R* (along with every other possible principle) is nonrejectable. He might argue that this is because, in Gremlin World, all possible principles are *tied* for having the weakest, strongest complaint against them—though it is a strong one indeed: that the world will be destroyed! I do not see how this reply can succeed, though. Unlike principle contractualism, Murphy's preferred version of contractualist moral reasoning begins its assessment of a general principle with a conversation between two people in which they ask for and give each other reasons. And it is certainly false that all possible principles, even in Gremlin World, are on a par with respect to *that* conversation. Plausibly, the reasons in favor of *R* are stronger than the reasons for a principle that permits what *R* forbids.

Murphy's problem, then, is that he "universalizes" too early, so to speak. As I have argued, it is better to first arrive at principles for the regulation of *my* behavior and then universalize to general principles. And maxim contractualism does just that.

VI

But now let us consider three objections to maxim contractualism. First, one may object that, by recommending that contractualists not commence their moral reasoning with the consideration of general moral principles, I am imploring contractualists to distance themselves from what, historically, has seemed like a deep insight—that morality is the product of an agreement among those whom it governs. How, this objection runs, can maxim contractualism model this insight since it requires no explicit appeal to a group of people deliberating but rather involves the consideration of a series of interactions between an agent and possible objectors?

This would be a problem for maxim contractualism because it would make it hard to see how regulating one's behavior in terms of the requirements derived in maxim contractualist moral reasoning demonstrates *respect* toward others—how complying with them would be part and parcel of treating others as *free and equal*. Contractualism, in its classical form, envisions morality not *merely* as the product of an agreement among those whom it governs but also as the product of an agreement among those whom it governs when they see each other as having a particular kind of standing. Pamela Hieronymi helpfully puts the point as follows:

> The principles of morality, as Scanlon understands it, are the principles that we would agree to in this contractualist situation. They are thus the terms of self-governance adopted by those who recognize each other as having a symmetric standing to determine the terms of their mutual self-governance. They are, we might say, the principles that would be agreed to in a Kingdom of Equals, each of whom is committed to living in a kind of harmony with the rest and so accords to each one a symmetric standing in determining the terms of his or her own self-governance.[43]

It is precisely because the principles of morality are those principles that people with equal standing would agree to that complying with them amounts to recognizing the equal standing of others (and violating them amounts to a failure of recognition).

43 Hieronymi, "Of Metaethics and Motivation," 106.

Despite appearances, I think this objection is mistaken. For, maxim contractualism *does* see morality as being the product of an agreement among those whom it governs. It is just that it takes the objects of agreement to be *personal* rather than general principles. Promisees, for example, could not agree to a principle that permitted *me* to break those promises I have made to them. Moreover, maxim contractualism understands the agreement about personal principles, which it models to be one had between those with equal standing. To stick with the previous example, *we* could not agree to a principle permitting me to break promises I have made, even if I could not gain from cooperation with the promisees in question. So, compliance with requirements derived in maxim contractualist moral reasoning does show respect to others. For the same reason, violation of them shows *disrespect*.

Second, it may be thought that principle contractualism is better placed to justify prohibitions against free-riding than maxim contractualism is and that this spells trouble for maxim contractualism. Here is how my objector may argue that principle contractualism is better placed than maxim contractualism is to justify prohibitions against free-riding: suppose you reside in a community in which it is customary not to litter. As a result of this convention, you and everyone else in the community benefit from there being clean streets. It seems like you are obligated not to litter yourself. To do so would be to free-ride on the effort of those people whose attitudes toward littering are necessary for the existence of the no-littering convention in your community. Were you about to litter, we could ask you: What if everyone did that? If everyone did that, the streets would not be clean in your community anymore. This is a weighty moral objection.

Principle contractualism can capture the moral force of the "what if everyone did that" question quite well. This is because it determines what you are morally required to do by assessing the implications of everyone complying with principles that permit you to do something. If everyone complied with a principle that permitted one to free-ride on the efforts of others, then the no-littering convention would no longer be in force in your community. And if the no-littering convention were no longer in force in your community, then there would not be clean streets in your community. And someone could put forward an objection in contractualist moral reasoning to that effect which is stronger than your reason to want to throw trash wherever.

Maxim contractualism, however, does not seem to be able to recognize such a complaint. In determining what you are morally required to do, maxim contractualism holds fixed what everyone else does. Maxim contractualist moral reasoning, after all, involves imagining what would happen if you adopted over time a maxim that might be behind your action *while holding fixed what everyone*

else does. So, maxim contractualist moral reasoning cannot appeal to what would happen in a scenario where everyone else starts littering. And, if you adopted a maxim over time that involved free-riding on socially beneficial conventions, the streets would not necessarily become unclean. Your acceptance of a norm that forbids people from littering is not necessary for the streets to be clean, so long as enough others accept that norm (and, by hypothesis, they do).

Such an objection to maxim contractualism, however, mischaracterizes the decisive complaint in contractualist moral reasoning to a maxim which involves free-riding on socially beneficial conventions. The complaint is not that the streets will be unclean as a result of the no-littering convention falling apart. It is rather that the adoption, over time, of a maxim that involves free-riding would result in treating unfairly those whose attitudes toward littering are necessary for the existence of the no-littering convention. From the perspective of those who are making the sacrifices necessary to maintain the no-littering convention, it would not be fair for people to get all the goods without doing any of the hard work necessary to get them.[44] And being treated unfairly is something that those people have strong reason to avoid.[45] Maxim contractualism, then, is just as able to justify prohibitions on free-riding as principle contractualism is.

At this juncture, one might have the following concern: When assessing a maxim in contractualist moral reasoning, how can one appeal to an interest in being treated fairly? One's strong reason to want to be treated fairly is a moral reason. But is not contractualism supposed to reduce the moral to the nonmoral? Such a concern, however, is unfounded. Scanlon explicitly allows moral considerations of certain kinds—fairness being one example—to enter

44 Murphy, in "Nonlegislative Justification," seems to agree: "If your child wants to throw the paper out the car window onto generally unlittered streets, the question 'What if everyone did that' is meant to get them to consider the *unfairness* of free-riding on others' beneficial compliance with the no-littering principle" (263, emphasis added). It is important to note, though, that Murphy's discussion is not aimed at showing how a "nonlegislative" version of contractualism might justify prohibitions against free-riding. Rather, Murphy is arguing that "what if everyone did that?" questions are best asked in free-riding contexts as opposed to all contexts in which actions are evaluated for moral permissibility, as rule consequentialists, principle contractualists, and other "legislative" moral theorists think.

45 Many have argued that consequentialists should be rule consequentialists rather than act consequentialists precisely in order to capture prohibitions on free-riding. But, given that the decisive complaint to free-riding in contractualist moral reasoning centers around the unfair treatment of those whose attitudes toward littering are necessary for the existence of the no-littering convention, act, maxim, and principle contractualism are all equally well-placed to justify prohibitions on free-riding. For a nice discussion of the consequentialist dialectic concerning prohibitions on free-riding, see Greene and Levinstein, "Act Consequentialism without Free Rides," 88–116.

into contractualist moral reasoning. Although my imagined objector is correct that Scanlon's project is reductionist, she is wrong to insist that what Scanlon means to reduce are moral facts. Rather, what Scanlon means to reduce is the deontic to the nondeontic.[46]

The third objection claims that maxim contractualism is subject to an ideal world problem of its own—one where one's future times slices play the same role that other people play in the ideal world problem for principle contractualism. To see how the second objection may arise, consider the following case.

> *Headache*: On Monday, Patient has an excruciating headache. Though it will go away on its own in five hours, drug *D* will cure it immediately. However, as *D* is a very potent drug, if it is administered on Monday, drug *E* must be administered on Tuesday in order to counter its side effects; otherwise, Patient will die. Doctor can administer *D* on Monday, but she knows that if she does so, even though she will be able to administer *E* on Tuesday, because of her own laziness then, she will not. Doctor has two options on Monday: administer *D* to Patient or do not.[47]

Intuitively, my objector may push, Doctor is not morally obligated to administer *D* to Patient. In fact, she may continue, Doctor is morally obligated to *not* administer *D* to Patient. She may argue that this is because, given Doctor's laziness, Doctor's administering *D* to Patient on Monday would have disastrous consequences. Patient will die on Tuesday if Doctor administers *D* to Patient on Monday. But—and here is her punch line—maxim contractualism entails that Doctor *is* morally obligated to administer *D* to Patient. So, my objector concludes, maxim contractualism is false.

Here is how my objector may argue that maxim contractualism entails that Doctor is morally obligated to administer *D* to Patient: according to maxim contractualism, Doctor is required to administer *D* to Patient just in case someone could reasonably reject any maxim that involved Doctor's not administering D to Patient. So, let us look at Doctor's maxim: "When in circumstances like Headache, I will not administer *D* to Patient." This maxim is rejectable. For,

46 For Scanlon's discussion, see *What We Owe to Each Other*, 212. I do not, then, find what is known in the literature as the "explanatory circularity objection" to be successful against Scanlonian contractualism. According to the explanatory circularity objection, Scanlonian contractualism implicitly relies on our pre-theoretical moral convictions when assessing candidate moral principles. On my view, this is no objection to Scanlonian contractualism if those convictions do not have *deontic content*—i.e., if those moral convictions do not express the claim that some action is right or wrong. (For the canonical presentation of the explanatory circularity objection, see Hooker, "Contractualism, Spare Wheel, Aggregation," 58.)

47 I borrow this case from Graham, "An Argument for Objective Possibilism," 220.

there is an alternative maxim that answers to a sufficient degree the reasons that Doctor has for favoring Doctor's adoption of the original maxim over time but whose implications do not justify objections to it from Patient that are as serious as those justified by the original maxim's implications. That maxim is: "In circumstances *C* like Headache, I will give *D and E*." If Doctor adopts this maxim, Patient will neither die nor continue to endure the headache. So, my objector infers, any maxim that involved Doctor's not administering *D* to Patient is rejectable. So, my objector concludes, according to maxim contractualism, Doctor is morally obligated to administer *D* to Patient—a most unintuitive result.

Whether or not Doctor is morally obligated to administer *D* to Patient is hotly contested in the literature by possibilists and actualists.[48] The actualist's argument that Doctor is not morally obligated to administer *D* (and, moreover, is obligated not to administer *D*) is that Doctor's administering *D* to Patient on Monday would have disastrous consequences—namely, that Patient will die on Tuesday if Doctor administers *D* on Monday. On the other hand, the possibilist argues that Doctor is morally obligated to administer *D* to Patient because Doctor is morally obligated to do the best she can for Patient, and there is something Doctor can do—namely, administer *D* on Monday and then administer *E* on Tuesday—that would cure Patient of her headache without killing her.[49] My objector, then, to borrow the terminology just introduced, wields actualist intuitions against maxim contractualism. Where my objector goes wrong is in her view that maxim contractualist moral reasoning is inherently possibilistic. It is not, and it is not precisely because it is possible to fail to act in accordance with a maxim one has adopted (and for a maxim contractualist to recognize this in her moral reasoning). I will turn now to showing how maxim contractualism can be adopted in either an actualist or a possibilist guise.

Actualists think that whether Doctor will administer *E* to Patient partly determines Doctor's obligations in Headache. And this is why actualists think that Doctor must not give *D* to Patient. Possibilists, on the other hand, think that whether Doctor will administer *E* to Patient is irrelevant to the determination of Doctor's obligations in Headache. In maxim contractualist moral reasoning, this contrast can be expressed in a choice-point as to whether Doctor's weakness of will with respect to a maxim is relevant to that maxim's rejectability or not. Here is how this would work: remember that, in order to determine whether someone is morally required to do something, maxim contractualists look at worlds in which that person adopts a maxim that involves not doing

48 For a helpful overview of the debate, see Timmerman and Cohen, "Moral Obligations."

49 Graham, "An Argument for Objective Possibilism," 221.

that thing. Possibilist maxim contractualists, when imagining such a world, do not take into account any weakness of will that that person has with respect to acting from that maxim. When determining whether Doctor ought to administer *D*, then, possibilist maxim contractualists only look at worlds in which Doctor *always* acts on the following maxim over time: "In circumstances *C* like Headache, I will give *D* and *E*." This feature of possibilist maxim contractualism is responsible for their agreement with my objector that any maxim that involved Doctor's not administering *D* to Patient is rejectable. If Doctor were to adopt the maxim "In circumstances *C* like Headache, I will give *D* and *E*" *and always administer both D and E*, then Patient (and other patients) would have their headaches cured and be at no risk of death. Of course, however, if you are a possibilist maxim contractualist, that your view entails that Doctor must administer *D* to Patient will not seem like an *objection* at all. You would just think it is an expression of the fact that morality does not let individuals off the hook for their contingent, avoidable limitations.

Actualist maxim contractualists, on the other hand, will take such a result of the possibilist version of their view to be a serious problem, indeed. This is why, in order to avoid the conclusion that Doctor must administer *D* to Patient, they hold fixed the contingent limitations of Doctor when imagining what would happen if Doctor adopted the maxim "In circumstances *C* like Headache, I will give *D* and *E*" over time. Given Doctor's weakness of will, his adoption of this maxim would result in him giving *D but not also administering E*, resulting in Patient's (and other patients') death. Rather than delivering the wrong result about Headache as my objector presses, maxim contractualism provides a framework within which possibilists and actualists can debate about what Doctor's obligations are.[50]

VII

Contractualists, then, need to alter their moral reasoning. When determining whether or not I must perform some action, they should not imagine what would happen if everyone did what I am considering doing. Instead, they

50 Much like with justifying prohibitions on free-riding, act, maxim, and principle contractualism are all equally well-placed to provide a framework within which possibilists and actualists can debate about what Doctor's obligations are. If you are a possibilist principle contractualist, then you will take Doctor's weakness of will with respect to a principle for the general regulation of behavior to be relevant to that principle's rejectability. If you are an actualist principle contractualist, then you will not. Similarly, if you are a possibilist act contractualist, then you will take Doctor's weakness of will with respect to administering *E* to Patient to be relevant to the rejectability of Doctor's administering *D* to Patient. And, if you are an actualist act contractualist, you will not.

should imagine what would happen if I adopted over time any maxim that might be behind my action while holding fixed what everyone else does. Such an alteration would allow contractualists to solve their ideal world problem while still providing them with the materials to vindicate our concept of morality as a social institution.[51]

Lingnan University
aaronsalomon@ln.edu.hk

REFERENCES

Ashford, Elizabeth. "The Demandingness of Scanlon's Contractualism." *Ethics* 113, no. 2 (January 2003): 273–302.

Graham, Peter A. "An Argument for Objective Possibilism." *Ergo* 6, no. 8 (2019–2020).

Greene, Preston, and Benjamin A. Levinstein. "Act Consequentialism without Free Rides." *Philosophical Perspectives* 34, no. 1 (December 2020): 88–116.

Gressis, Rob. "Recent Work on Kantian Maxims 1: Established Approaches." *Philosophy Compass* 5, no. 3 (March 2010): 216–27.

Hieronymi, Pamela. "Of Metaethics and Motivation: The Appeal of Contractualism." In *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, edited by R. Jay Wallace, Rahul Kumar, and Samuel Freeman, 101–28. New York: Oxford University Press, 2011.

Hills, Alison. "Utilitarianism, Contractualism, and Demandingness." *Philosophical Quarterly* 60, no. 239 (April 2010): 225–42.

Hooker, Brad. "Contractualism, Spare Wheel, Aggregation." In *Scanlon and Contractualism*, edited by Matt Matravers, 53–76. London: Frank Cass, 2003.

Kagan, Shelly. "Evaluative Focal Points." In *Morality, Rules, and Consequences: A Critical Reader*, edited by Brad Hooker, Elinor Mason, and Dale Miller, 134–55. Edinburgh: Rowman & Littlefield, 2000.

———. *The Limits of Morality*. New York: Oxford University Press, 1989.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. Edited and translated by Mary Gregor and Jen Timmerman. New York: Cambridge University Press, 2012.

Korsgaard, Christine M. "Acting for a Reason." In *The Constitution of Agency:*

51 I am grateful for discussion with and written comments from Brian Berkey, Jed Lewinsohn, Andrew Lichter, Anthony Nguyen, Japa Pallikkathayil, Kyra Salomon, Aaron Segal, Anna-Bella Sicilia, Anna Stilz, Michael Thompson, Daniel Webber, and Pablo Zendejas Medina. Thanks also to two anonymous referees from the *Journal of Ethics and Social Philosophy*.

Essays on Practical Reason and Moral Psychology, 207–30. New York: Oxford University Press, 2008.

———. "Morality as Freedom." In *Creating the Kingdom of Ends*, 159–87. Cambridge: Cambridge University Press, 1996.

Kumar, Rahul. "Defending the Moral Moderate: Contractualism and Common Sense." *Philosophy and Public Affairs* 28, no. 4 (October 1999): 275–309.

Murphy, Liam. "Nonlegislative Justification." In *Principles and Persons: The Legacy of Derek Parfit*, edited by Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan, 247–76. New York: Oxford University Press, 2021.

O'Neill, Onora. "A Simplified Account of Kant's Ethics." In *Contemporary Moral Problems*, edited by James E. White. St. Paul, MN: West Publishing Co., 1985.

Parfit, Derek. *On What Matters*. 2 vols. New York: Oxford University Press, 2011.

Podgorski, Abelard. "Wouldn't It Be Nice? Moral Rules and Distant Worlds." *Noûs* 52, no. 2 (June 2018): 279–94.

Potter, Nelson T., and Mark Timmons. "Introduction." In *Morality and Universality: Essays on Ethical Universalizability*, edited by Nelson T. Potter and Mark Timmons, ix–xxxii. Dordrecht: D. Reidel Publishing Company, 1985.

Rabinowicz, Wlodzimierz. *Universalizability: A Study in Morals and Metaphysics*. Dordrecht: D. Reidel Publishing Company, 1979.

Rawls, John. "Two Concepts of Rules." *Philosophical Review* 64, no. 1 (January 1955): 3–32.

Rosen, Gideon. "Might Kantian Contractualism Be the Supreme Principle of Morality?" *Ratio* 22, no. 1 (March 2009): 78–97.

Scanlon, T. M. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, MA: Harvard University Press, 2008.

———. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press, 1998.

Sheinman, Hanoch. "Act and Principle Contractualism." *Utilitas* 23, no. 3 (September 2011): 288–315.

Thomson, Judith Jarvis. "Self-Defense." *Philosophy and Public Affairs* 20, no. 4 (Autumn 1991): 283–310.

Timmerman, Travis, and Yishai Cohen. "Moral Obligations: Actualist, Possibilist, or Hybridist?" *Australasian Journal of Philosophy* 94, no. 4 (February 2016): 672–86.

Walden, Kenneth. "Mores and Morals: Metaethics and the Social World." In *The Routledge Handbook of Metaethics*, edited by Tristram McPherson and David Plunkett, 417–31. New York: Routledge, 2017.

Journal of Ethics and Social Philosophy
VOL. 26, NO. 3 · FEBRUARY 2024
https://doi.org/10.26556/jesp.v26i3.2103

NONNATURALISM, THE SUPERVENIENCE CHALLENGE, HIGHER-ORDER PROPERTIES, AND TROPE THEORY

Jussi Suikkanen

According to nonnaturalist realism, normative properties are unique kinds of stance-independent properties.[1] However, many metaethicists reject this view because of the *supervenience challenge*: the nonnaturalists arguably fail to explain why two otherwise identical actions cannot have different normative properties. Section 1 below outlines nonnaturalist realism and the supervenience challenge in more detail.

Mark Schroeder and Knut Olav Skarsaune have recently introduced an elegant nonnaturalist response to this challenge.[2] They suggest that nonnaturalists should take *action kinds* to be the primary bearers of normative properties.[3] The ascriptions of those properties to action tokens should then be understood to be about these tokens belonging to the kinds that instantiate the normative properties. Because two tokens that share the same base properties belong to the same kinds, the supervenience of the normative properties on the natural properties seems to follow, as section 3 explains.

This article develops the previous response in two ways. First, it gives additional support for Schroeder's and Skarsaune's thesis that normative properties are primarily instantiated by action kinds. Hence, section 2 explains two arguments for that thesis based on the work of H. A. Prichard.[4]

1 See, e.g., Shafer-Landau, *Moral Realism*; Huemer, *Ethical Intuitionism*; Fitzpatrick, "Robust Ethical Realism, Nonnaturalism, and Normativity"; Cuneo, *The Normative Web*; Enoch, *Taking Morality Seriously*; and Wielenberg, *Robust Ethics*.

2 Schroeder, "The Price of Supervenience"; Skarsaune, "How to Be a Moral Platonist." For a resembling strategy, see Scanlon, *Being Realistic about Reasons*, ch. 2. For objections to Scanlon's formulation of the response, see Schroeder, "The Price of Supervenience," 136–37. Skarsaune himself got at least a part of the idea from Kit Fine in discussion ("How to Be a Moral Platonist," 245n1).

3 Depending on the normative property, we could equally take the bearers to be outcome kinds, character kinds, and so on. For simplicity's sake, I focus on action kinds.

4 Prichard, "Duty and Ignorance of Fact."

Second, both Schroeder and Skarsaune recognize that their response works *only if* action kinds have their normative properties necessarily (section 3). In response to this problem, Skarsaune relies on transcendent realism about universals.[5] Section 4 argues that this proposal is problematic both (i) dialectically, as the defenders of the supervenience challenge will object to the additional metaphysical commitments the proposal requires, and (ii) because there are well-known reasons to reject transcendent realism about universals. Finally, section 5 develops Schroeder's and Skarsaune's response further in the framework of contemporary trope theory. This enables the nonnaturalist realists to respond to the supervenience challenge by relying on a plausible mainstream view of properties, the adoption of which does not require any further metaphysical commitments beyond the nonnatural properties themselves.

1. NONNATURALISM AND THE SUPERVENIENCE CHALLENGE

Nonnaturalist realism consists of the following theses:

> PROPERTIES: There are normative properties, and these properties are instantiated in the actual world.
>
> INDEPENDENCE: Normative properties are stance independent.
>
> DISTINCTNESS: Normative properties are of their own unique kind.[6]

PROPERTIES rules out error theory (the view that normative are not instantiated), expressivism, and quietism. According to the latter views, we can talk about normative properties and their instantiation, but such talk is to be understood in a deflationary way.[7] In contrast, nonnaturalist realists are committed to the existence of metaphysically robust normative properties that can do explanatory work.

INDEPENDENCE rules out constructivism, contextualism, constitutivism, relativism, subjectivism, and response-dependence theories. According to them, normative properties are grounded in the attitudes and judgments of

5 Skarsaune, "How to Be a Moral Platonist," sec. 10.7.

6 For similar understandings of nonnaturalist realism, see, e.g., McPherson, "Ethical Nonnaturalism and the Metaphysics of Supervenience," 207–10; Dreier, "Explaining the Quasi-Real," 277, and "Is There a Supervenience Problem for Robust Moral Realism?," 1392–93; Leary, "Nonnaturalism and Normative Necessities," 78–79; Toppinen, "Nonnaturalism Gone Quasi," 25; and Väyrynen, "The Supervenience Challenge to Nonnaturalism," 171.

7 Blackburn, *Essays in Quasi-Realism*; Scanlon, *Being Realistic about Reasons*.

either actual or idealized agents.[8] In contrast, the central *realist* claim of nonnaturalism is that normative properties are "stance independent."[9]

DISTINCTNESS finally rules out different forms of naturalism. It states that normative properties are not themselves irreducible natural properties nor reducible to such properties, but rather they are wholly different kinds of properties.[10] This *noncontinuity thesis* requires that we can characterize the distinguishing features of natural properties. The most promising suggestions are that they are the subject matter of natural sciences, invoked in scientific explanations, known *a posteriori*, causally efficacious, and/or figure in the laws of nature.[11] Nonnaturalists thus claim that the normative properties lack the previous features.

The following then captures how normative properties are thought to supervene on the base properties:

> SUPERVENIENCE: It is conceptually necessary that when something has a normative property *N*, it also has a base property *P* such that it is metaphysically necessary that anything else that is *P* also is *N*.[12]

SUPERVENIENCE refers to two kinds of properties and necessities. The normative property in it can be any normative property we ascribe to actions with normative predicates ("ought," "good," and the like). As we saw, nonnaturalists claim that these properties are unique kinds of properties, but SUPERVENIENCE itself is neutral about their metaphysical nature.

SUPERVENIENCE also mentions a base property *P*. It is neither a *sui generis* normative property nor a property the correct analysis of which ineliminably

8 Dunaway, "Epistemological Motivations for Anti-Realism," sec. 1.

9 Shafer-Landau, *Moral Realism*, 49.

10 Thus, according to DISTINCTNESS, normative propositions are not entailed by propositions that ascribe natural properties. See Blackburn, *Essays in Quasi-Realism*, 116; and Skarsaune, "How to Be a Moral Platonist," 247.

11 See, Moore, *Principia Ethica*, 40; Little, "Moral Realism II," 26; Copp, "Why Naturalism?"; Lewis, "New Work for a Theory of Universals"; and Vallentyne, "The Nomic Role Account of Carving Reality at the Joints."

12 See Dreier, "The Supervenience Argument against Moral Realism," 14–17, "Explaining the Quasi-Real," 275, and "Is There a Supervenience Problem for Robust Moral Realism?," 1393; Wedgwood, "The Price of Non-Reductive Moral Realism," sec. 2; Skarsaune, "How to Be a Moral Platonist," 247–48; Leary, "Nonnaturalism and Normative Necessities," 80; Toppinen, "Nonnaturalism Gone Quasi," 28; and Väyrynen, "The Supervenience Challenge to Nonnaturalism," 172–73. McPherson formulates a different, global supervenience claim for the purposes of the challenge ("Ethical Nonnaturalism and the Metaphysics of Supervenience," 210–17), whereas Schroeder outlines the challenge without the previous kind of full-fledged supervenience ("The Price of Supervenience," 126).

mentions normative properties, i.e., not a normative property as understood by the nonnaturalists.[13] Roughly, we can take *P* to be a factual, natural, nonnormative property, though these characterizations are not metaethically neutral.[14] *P* can also be complex: a conjunctive property of having p_1, p_2, …, and p_n where these properties are intrinsic and relational nonnormative properties of a given action. SUPERVENIENCE then claims that a part of the meaning of normative concepts is that, when something has a normative property, it also has a base property, the having of which *metaphysically necessitates* having that normative property.

Consider Ann, who helps an elderly person across the road, and the conjunction of all the nonnormative properties of this action.[15] This property includes all the nonnormative features of the action, including Ann's motivations and the action's consequences. Intuitively we also think that Ann did something good. Image then that Ben accepts this but goes on to describe another action exactly like Ann's, which has all the same base properties and only those. Ben then, however, claims that even if Ann does something good, the other action is not good at all. Here we would think that Ben is confused, incompetent with the normative terminology. Ben cannot, for example, describe what makes Ann's action good and the second action not good, given that both actions are otherwise identical. Of course, Ann's action could have been not good too, but only in the sense that, if that action had been different, it would have been not good.

Cases like this illuminate and support SUPERVENIENCE. They suggest that two actions cannot have different normative properties unless they differ in some more basic respect, and they also illustrate the idea that the metaphysically necessary connection between the two different kinds of properties is required by conceptual necessity.

We then have all the elements of the supervenience challenge.[16] The nonnaturalist realists claim that the supervening normative properties and the

13 McPherson, "Ethical Nonnaturalism and the Metaphysics of Supervenience," 213–14.

14 McPherson, "Ethical Nonnaturalism and the Metaphysics of Supervenience," 214–15, and "Supervenience in Ethics," sec. 1.1; Sturgeon, "Doubts about the Supervenience of the Evaluative."

15 This example draws from Hare, *The Language of Morals*, 81; and Dreier, "The Supervenience Argument against Moral Realism," 16, "Explaining the Quasi-Real," 276, and "Is There a Supervenience Problem for Robust Realism?," 1395–96. See also Blackburn, *Essays in Quasi-Realism*, 116.

16 See Hare, *The Language of Morals*, sec. 10.2; and Blackburn, *Essays in Quasi-Realism*, essays 6 (esp. pp. 118–19) and 7. For the historical development of the argument, see Dreier, "Is There a Supervenience Problem for Robust Moral Realism?," sec. 1.2. My presentation follows Schroeder's second way of formulating the challenge ("The Price of Supervenience," 127–28). See also Dreier, "The Supervenience Argument against Moral Realism," 16–17, and "Explaining the Quasi-Real," 274–76; Gibbard, *Thinking How to Live*, 20; Ridge,

nonnormative base properties are *discontinuous*. If SUPERVENIENCE is true, they also must grant that there is a metaphysically necessary connection between these discontinuous properties.

There are necessary connections between seemingly discontinuous properties elsewhere too. For example, it is metaphysically necessary that anything that is hot has the property of having high average kinetic energy of particles. Here, however, we have an explanation of the necessary connection: the former property is reducible to the latter. Philosophers, thus, generally tend to explain necessary connections between seemingly discontinuous properties by showing that one of the properties can be (i) analyzed in terms of, (ii) reduced to, or (iii) identified with the other property. However, the nonnaturalist realists cannot rely on these explanations because for them the normative properties are *sui generis* in a way that blocks analyses, reductions, and identities.[17]

The threat, then, is that the nonnaturalist realists must grant that the necessary metaphysical connection in question is *brute*—a connection that cannot be explained.[18] Yet, an attractive methodological principle is that a "commitment to brute necessary connections between discontinuous properties counts significantly against a view."[19] The supervenience challenge, then, is that the nonnaturalist realists must provide an explanation of the necessary metaphysical connection between the normative properties and their base properties that is compatible with the normative properties being discontinuous, or otherwise, we have good reasons to prefer other metaethical views that can avoid similar brute connections.[20]

"Anti-Reductionism and Supervenience," sec. 1; McPherson, "Ethical Nonnaturalism and the Metaphysics of Supervenience," sec. 3, and "Supervenience in Ethics," sec. 4; Skarsaune, "How to Be a Moral Platonist," 249–50, 266; Leary, "Nonnaturalism and Normative Necessities," 80–81; Toppinen, "Nonnaturalism Gone Quasi," 28; and Väyrynen, "The Supervenience Challenge to Nonnaturalism," 174. For a general discussion, see Van Cleve, "Brute Necessity."

17 This is why there is no supervenience challenge for the naturalist versions of realism as natural properties supervene trivially on the natural base properties (Dreier, "The Supervenience Argument against Moral Realism," and "Explaining the Quasi-Real," 277; McPherson, "Supervenience in Ethics," secs. 4.1–4.2). Expressivists have also argued that they do not face the challenge (Hare, *The Language of Morals*, 14; Blackburn, *Essays in Quasi-Realism*, 122, 137; Gibbard, *Thinking How to Live*, 90–98), but this is challenged by Dreier ("Explaining the Quasi-Real"). For a response, see Toppinen, "Nonnaturalism Gone Quasi."

18 Gibbard, *Thinking How to Live*, 20.

19 McPherson, "Ethical Nonnaturalism and the Metaphysics of Supervenience," 217–18.

20 Some philosophers, such as Shoemaker ("Causality and Properties") and Swoyer ("The Nature of Natural Laws"), reject the previous Humean assumption and argue for brute metaphysical necessary connections between properties. If they are correct, the supervenience challenge for the nonnaturalist realists collapses and needs no answer. I merely

It is important to emphasize here that SUPERVENIENCE contains two necessities: one conceptual and one metaphysical. The supervenience challenge is to explain the second—metaphysical—necessity. The first conceptual necessity tells us only that if there are normative properties, they must be metaphysically necessitated by the base properties. This conceptual truth calls for a conceptual explanation, but those conceptual explanations will be metaethically neutral.[21] They are even compatible with error theory—the view that there are no normative properties that are so related to the base properties. What the nonnaturalists must then explain is the second metaphysical necessity—that is, how there can be normative properties that are related to the base properties as the conceptual truth requires them to be connected. For this reason, I focus below solely on explaining the second metaphysical necessity.[22]

2. NORMATIVE PROPERTIES AND KINDS

There are many nonnaturalist attempts to respond to the previous challenge.[23] This article explores Schroeder's and Skarsaune's suggestion, according to

argue that, even if the previous philosophers were mistaken and the Humean assumption were a reasonable methodological principle, the supervenience challenge could still be responded to.

21 Stratton-Lake and Hooker, "Scanlon vs. Moore on Goodness."

22 See Ridge, "Anti-Reductionism and Supervenience"; McPherson, "Ethical Nonnaturalism and the Metaphysics of Supervenience"; and Dreier, "Is There a Supervenience Problem for Robust Moral Realism?," sec. 2.4. Many philosophers think, however, that there are no conceptual truths. Furthermore, others might at least argue that the fact that the instantiation of the relevant base properties necessitates the instantiation of the normative properties is not a conceptual truth even if there are others. Some of these philosophers might still think that the previous metaphysical necessitation relation both holds and calls for an explanation. For this reason, in responding to the supervenience challenge, we can remain neutral about the previous conceptual truth as long as we believe that the metaphysical necessitation relation holds.

23 It has been argued that (i) supervenience is a moral doctrine rather than a metaphysical or a conceptual claim in need of an explanation (Kramer, "Supervenience as an Ethical Phenomenon"; for objections, see Dreier, "Explaining the Quasi-Real," 278, and "Is There a Supervenience Problem for Robust Moral Realism?," sec. 2.3; McPherson, "Ethical Nonnaturalism and the Metaphysics of Supervenience," 220–21, and "Supervenience in Ethics," sec. 4.5); that (ii) a conceptual explanation of the supervenience is sufficient and so the nonnaturalists do not have to provide a metaphysical explanation (Shafer-Landau, *Moral Realism*, 86; Stratton-Lake and Hooker, "Scanlon vs. Moore on Goodness," 164; Enoch, *Taking Morality Seriously*, 149; Olson, *Moral Error Theory*, 90, 96; for objections, see McPherson, "Ethical Nonnaturalism and the Metaphysics of Supervenience," 221–2, and "Supervenience in Ethics," sec. 4.4; Dreier, "Explaining the Quasi-Real," 281, and "Is There a Supervenience Problem for Robust Moral Realism?," sec. 2.4; and Väyrynen,

which nonnaturalist realists can respond to the challenge by claiming that normative properties are primarily instantiated by action kinds rather than action tokens.[24] I will explain how this idea helps with the supervenience challenge in section 3, but before that, this section provides additional support for one important element of the response.

One essential part of Schroeder's and Skarnause's response is the conditional claim that "if the normative properties are primarily instantiated by action kinds, the nonnaturalist realists can respond to the supervenience challenge." Sections 3–6 below will focus on arguing for this claim. This claim is also the main focus of this article as it is important for the nonnaturalist realists to specify the conditions under which the supervenience challenge can be met—it can be met as long as action kinds are the primary bearers of normative properties, or so I will argue below. Yet, before that, I believe that this nonnaturalist realist's response to the supervenience challenge is even stronger the more plausible the antecedent of the previous conditional can be made: the more reasons can be given for thinking that action kinds, in fact, are the primary bearers of normative properties. In this case, we would not only know under which conditions the supervenience challenge would be met, but we would also have good reasons to think that those conditions are actually met. Thus, the aim of this section is to make Schroeder's and Skarsaune's response

"The Supervenience Challenge to Nonnaturalism," 175–76); that (iii) a metaphysical "making-relation" to be captured in the fundamental normative laws is sufficient to provide the explanation (Enoch, *Taking Morality Seriously*, ch. 6; Scanlon, *Being Realistic about Reasons*, 40–41; Olson, *Moral Error Theory*, 97–100; Wielenberg, *Robust Ethics*, ch. 1; for objections, see Toppinen, "Nonnaturalism Gone Quasi," 29; and Leary, "Nonnaturalism and Normative Necessities," 87); that (iv) the normative facts are exhaustively constituted by nonnormative facts (Shafer-Landau, *Moral Realism*, 87–88; for an objection, see McPherson, "Ethical Nonnaturalism and the Metaphysics of Supervenience," 226; Leary, "Nonnaturalism and Normative Necessities," 89–93; and Väyrynen, "The Supervenience Challenge to Nonnaturalism," 176–77); that (v) we should reject SUPERVENIENCE and so there is nothing for the nonnaturalist realist to explain (Fine, "Varieties of Necessity"; Rosen, "Metaphysical Relations in Metaethics"; for objections, see McPherson, "Supervenience in Ethics," sec. 4.3; Väyrynen, "The Supervenience Challenge to Nonnaturalism," 180–82; and Dreier, "Is There a Supervenience Problem for Robust Moral Realism?," sec. 2.5); that (vi) supervenience can be explained by relying on the essences of normative properties (Leary, "Nonnaturalism and Normative Necessities"; for an objection see McPherson, "Ethical Nonnaturalism and the Metaphysics of Supervenience," 223); and that (vii) the objection relies on flawed principles of modal logic (Wedgwood, "The Price of Non-Reductive Moral Realism"; for an objection see Schmitt and Schroeder, "Supervenience Arguments under Relaxed Assumptions").

24 Schroeder, "The Price of Supervenience"; Skarsaune, "How to Be a Moral Platonist."

stronger by providing additional support for the claim that the primary bearers of normative properties are action kinds.[25]

The claim that normative properties are primarily instantiated by action kinds was first put forward by H. A. Prichard, which is why it has become known as "the Prichard point."[26] There are two independent kinds of support for this claim. The first relies on our intuitions about what we ought to do, whereas the second relies on an argument first put forward by Prichard himself.

To get a sense of the first, intuitive type of support, let us focus on *ought* as a paradigmatic normative property. Here is an intuitive reason to think that this property is primarily instantiated by action kinds.[27]

If I owe you five dollars, I ought to pay you the money back when you ask for it. Yet, consider the different ways in which I could do so: either today or tomorrow, in cash or by check, graciously or churlishly, here or there … The intuitive thought is that, taken individually, none of these specific action tokens has the property of being what I ought to do. Rather, what has that property is the more general action kind—the kind to which most action tokens that consist of me paying you back belong.[28]

Prichard makes the second argument in the following passage:

> But, as we recognize when we reflect, there are no such characteristics of an action as ought-to-be-doneness and ought-not-to-be-doneness. This is obvious; for, since the existence of an obligation to do some action

25 The fact that taking normative properties to be primarily instantiated by action types helps with the supervenience challenge already provides some reason to think that those properties really are primarily instantiated by action kinds as problem-solving and explanatory powers are arguably one reason to accept metaphysical claims such as this (see also Schroeder, "The Price of Supervenience," 141; and Skarsaune, "How to Be a Moral Platonist," 255). Skarsaune, in addition, refers to more general linguistic evidence and furthermore argues that the direction of epistemic justification usually proceeds from general normative judgments about action kinds and empirical information to normative judgments about cases ("How to Be a Moral Platonist," sec. 10.3 and pp. 260–62). For a similar argument based on Price (*A Review of the Principal Questions in Morals*) and Cudworth (*A Treatise Concerning Eternal and Immutable Morality*), see Schroeder, "The Price of Supervenience," 138–40.

26 Prichard, "Duty and Ignorance of Fact," 98–100; Dancy, *Practical Shape*, 30–33. In addition to Prichard and Dancy, other notable defenders of this claim include Anscombe (*Intention*, 6) and Stocker ("Duty and Supererogation," 54).

27 Dancy, *Practical Shape*, 31; Stocker, "Duty and Supererogation," 54.

28 This argument, admittedly, relies on ought being "uniqueness entailing," on the idea that in any situation at most one action is what you ought to do. Other normative properties, such as goodness, permissibility, or kindness, are not like this. It can be good both to pay your friend in cash and to pay them with a check. This is why this argument does not directly support the general conclusion that all normative properties are primarily instantiated by kinds.

> cannot possibly depend on actual performance of the action, the obligation itself cannot be a property which the action would have, if it were done.[29]

According to Dancy, Prichard begins from assuming that we consider normative properties when we are thinking about what we are to do in the future.[30] He also needs another assumption—namely that, if there is something we ought to do in the future, we are already now under an obligation to act in that way then, even if that action has not yet been done.

Prichard's argument then is that, from the temporally antecedent perspective, an action token that does not yet exist but merely could exist in the future cannot now have the property of being what I am actually already now obliged to do in the future. This is because no such action token exists now, before I have done the action, to have that property. It could, of course, be suggested that an action token, when it becomes actual in the future when I do the action, will then have the property of being what I ought to do. However, as Prichard points out, this is not enough: it does not oblige me now that an action would in the future (if I were to do the action) be such that I ought to do it then.[31]

It is more plausible to focus from the previous antecedent perspective on just *how* we ought to act. When we do so, we are explicitly focusing on kinds of actions and what normative properties they have. To make sense of what we are actually obliged to do and of how we deliberate, we thus ought to think that normative properties are primarily instantiated by action kinds. One way to put this is that when I decide to pay back the money I owe to you, this is not choosing an action token like I would choose a chocolate from a box. Rather, it is just to decide that one of my future actions will be of a certain reimbursing kind, which is a kind of action I ought to do.[32]

There are then reasons to think that normative properties are primarily instantiated by action kinds. However, before we move on, I need to introduce a piece of terminology to make things simpler. Following Skarsaune, I will take being a certain kind of an action to be a first-order property, which an action token can have.[33] This enables us to understand normative properties,

29 Prichard, "Duty and Ignorance of Fact," 99.

30 Dancy, *Practical Shape*, 31.

31 This argument assumes that particular actions are concrete events, as most philosophers believe they are. An alternative view is that they are ordered triplets of agents, action types, and times (Goldman, *A Theory of Human Action*). On such a view, particular actions may exist before they are performed, but presumably, their normative properties would naturally, in this case, be instantiated by the action types that in part constitute the particular action.

32 Dancy, *Practical Shape*, 32.

33 Skarsaune, "How to Be a Moral Platonist," 268.

given that they are primarily instantiated by action kinds (i.e., by the first-order properties) as higher-order properties.

3. THE HIGHER-ORDER PROPERTY SOLUTION TO THE SUPERVENIENCE CHALLENGE

We can then state the higher-order property solution to the supervenience challenge. Section 2 argued that normative properties are primarily instantiated by the first-order properties of action tokens belonging to certain action kinds. Yet, in ordinary language, we also ascribe normative properties to action tokens. We might say:

1. Ben should not have said that.

The key to understanding Schroeder and Skarsaune is to begin from how they suggest the nonnaturalists should analyze this claim, assuming that normative properties are primarily instantiated by action kinds.

The suggestion is that we should understand claims like 1 as "mixed" (or, in Schroeder's terms, "bastard") normative claims that are to be reductively analyzed in the following way:

$$\exists K\,(\text{token}(\text{Ben's utterance}, K)\ \&\ \text{should-not}_{\text{kind}}(K)).^{34}$$

This reductive analysis states that the truth conditions of 1 consist of there being some action kind to which Ben's utterance belongs such that that kind of action has the normative property of being what one should not do. Utterance 1 is thus analyzed in terms of (i) the action token belonging to a kind (i.e., the empirical, contingent part) and (ii) that kind instantiating the relevant normative property (i.e., the pure normative part).[35]

34 Skarsaune, "How to Be a Moral Platonist," 252, 260. See also Schroeder, "The Price of Supervenience," 131, 141. Here the variable *K* ranges over descriptive kinds (of events, actions, outcomes, etc.) but not over haecceitic kinds (Skarsaune, "How to Be a Moral Platonist," 252–53). For evidence for the claim that we should analyze normative claims about action tokens in this way rather than, in the other direction, normative claims about action types in terms of truths about tokens (that is, for the ascriptions of normative properties to kinds to be more fundamental than the ascriptions of normative properties to tokens), see Skarsaune, "How to Be a Moral Platonist," sec. 10.4, as well as section 2 above.

35 Structurally, this analysis of claims that ascribe normative properties to tokens is similar to Hare's universal prescriptivism (*The Language of Morals*) and Gibbard's norm-expressivism (*Wise Choices, Apt Feeling*) as they analyze normative predications to a particular in terms of general commitments (see Skarsaune, "How to Be a Moral Platonist," 254).

How does this help with the supervenience challenge? Let us assume that Charlie makes an utterance that has exactly the same nonnormative base properties as Ben's. Consider the following utterance:

2. Charlie should not have said that.

The model above reductively analyzes this claim as follows:

$\exists K$ (token(Charlie's utterance, K) & should-not$_{\text{kind}}$(K)).

Let us assume that 1 is true. The second conjunct in both analyses of 1 and 2 is the same—that there is a kind K that has the property of being what one should not do. So, if we assume that this element is true in one case, it should be true in the other case, too (though see below). The first part of the analysis of 1 states that Ben's utterance belongs to the kind K that has the relevant normative property. However, if Charlie's utterance has exactly the same nonnormative base properties, it must belong to the same action kind as Ben's utterance. After all, if two actions have the same nonnormative base properties, they cannot belong to different kinds.[36]

Of course, if the two action tokens had different base nonnormative properties, they could belong to different kinds, one which could have the relevant normative property and the other lack it. The supervenience thesis, however, only requires that there cannot be a normative difference without a difference in the nonnormative base properties, and so we have an explanation of why the normative supervenes on the natural.[37]

There is, however, a gap in this response, which both Schroeder and Skarsaune recognize.[38] It does not work if it is possible both (i) that Ben is in a possible world in which the relevant kind K to which his utterance belongs has the property of being what you should do and (ii) that Charlie is simultaneously in a different world in which K does not have that normative property. If that is a possibility, then even if Ben's and Charlie's utterances had the same base properties, 1 could be true and 2 false, and so the response would fail.

Both Schroeder and Skarsaune thus recognize that their suggestion works *only if* the relevant action kinds have their normative properties *necessarily*—always and across all possible worlds. Furthermore, the nonnaturalist realists

36 About this brute connection and the explanation for it, see Skarsaune, "How to Be a Moral Platonist," 267.

37 Schroeder, "The Price of Supervenience," 132; Skarsaune, "How to Be a Moral Platonist," 267. There is still a question of which action kinds have which normative properties. The nonnaturalists must take this connection to be brute (Schroeder, "The Price of Supervenience," 144; Skarsaune, "How to Be a Moral Platonist," 268–69).

38 Schroeder, "The Price of Supervenience," 142; Skarsaune, "How to Be a Moral Platonist," 263.

cannot just insist that this is the case as this would commit them to brute necessary connections between distinct existences. This is why it might look like no progress has been made.

4. THE SUPERVENIENCE CHALLENGE AND TRANSCENDENT UNIVERSALS

Skarsaune addresses the previous problem in the following way.[39] Using the terminology of section 2, we can take belonging to the action kinds that instantiate the relevant normative properties to be first-order properties of action tokens. We can then understand normative properties as second-order properties instantiated by the previous first-order properties.

The first part of Skarsaune's proposal is that the nonnaturalist realists should take the previous two properties to be *universals*.[40] The fact that a certain action kind has a certain normative property can then be understood in terms of an instantiation relation between the two properties: the first-order property, as a universal, of being of a certain kind of an action instantiates a higher-order property, another universal, of being, say, wrong.

The second part is that the nonnaturalist realists should then adopt moral Platonism based on transcendent realism about universals.[41] On this view, the fact that one universal instantiates another is not a fact that obtains in virtue of what is the case in this or that possible world or even within all worlds. Rather, such a fact is "transcendent"—one that obtains independently of all worlds.[42] There is thus nothing in the possible worlds that makes it true that the action kind "helping others" has the higher-order property of goodness. Rather, the relevant universals are abstract entities, which exist in a transcendent realm outside space, time, and the possible worlds. They form an invariable framework of what can be the case within all possible worlds.[43]

This proposal provides a nonnaturalist response to the supervenience challenge. The initial gap in Schroeder's and Skarsaune's response was that nothing in it guaranteed that the relevant action kinds have their normative properties in all possibilities. However, the previous addition suggests that the fact that the instantiation relation between a first-order action kind universal and a second-order normative property universal obtains in the distinct transcendent

39 Skarsaune, "How to Be a Moral Platonist," sec. 10.7.

40 Skarsaune, "How to Be a Moral Platonist," 268.

41 Skarsaune, "How to Be a Moral Platonist," 270.

42 Fine, "Necessity and Nonexistence," 325–26; Skarsaune, "How to Be a Moral Platonist," 270.

43 For classic defenses, see Plato, *The Republic*, bk. 7; and Russell, "The World of Universals."

realm outside all possibilities explains why in all worlds, the relevant action kind instantiates the given normative property.

Here we do not need to add that the relevant action kind universal instantiates the normative property universal in the transcendent realm "necessarily." This is because, in this realm, there are no different cases or *possibilia* in which sometimes the former universal instantiates the latter universal and sometimes it does not. There is just one action kind universal, one normative property universal, and an instantiation relation between them in one atemporal and aspatial realm where everything is immutable and indestructible. For this reason, there cannot be different *cases* where there could be variation in whether the instantiation relation holds between the two universals. Furthermore, what is the case in the transcendent realm then determines how things are within all possible worlds. We thus get an explanation of the necessary connection between an action kind and a normative property in terms of how these universals are related in the transcendent realm.[44]

If the previous metaphysical picture is acceptable, the nonnaturalists have a response to the supervenience challenge. The problem, however, is that we have been given little reason to believe that the relevant first-order and second-order properties should be understood as transcendent universals. This leads to two problems. First, the solution is dialectically problematic, and second, there are well-known objections to transcendent realism about universals.

In terms of the dialectic, the supervenience challenge objection to nonnaturalism is usually made by those who have deep naturalist sympathies.[45] The

44 According to nonnaturalist realism, which action kinds instantiate specifically which normative property universals cannot be explained further. For why this is not a problem, see Skarsaune, "How to Be a Moral Platonist," 268–69. The inference above does not move from "immutable" and "indestructible" to "necessary" as this would be a fallacy (the date of my birthday is immutable and indestructible though not necessary). The key is that there is only one case of the instantiation relation between the universals that determines how things are in all worlds.

45 For the naturalist commitments of the key defenders of the objection, see, e.g., Blackburn, *Ruling Passions*, 48–49; Gibbard, *Reconciling Our Aims*, 14–17; and Dreier, "Another World," 158. It could be suggested that Hare had something like the objection in mind, even if he was a theist and thus a nonnaturalist (*The Language of Morals*). This would suggest that naturalist commitments are not essential to the objection itself. Here it is worthwhile to note that in his philosophy of religion, due to his empiricist and naturalist views of meaning, even Hare rejected transcendent God and so tried to find ways of understanding his own theism and religious beliefs and utterances in a way that would be compatible with his naturalism (Hare, "The Simple Believer"). It is true, however, that naturalism is not essential to the supervenience challenge as a person who is a nonnaturalist about something other than normative properties can object to nonnaturalism about normative properties on the basis of that challenge. Yet, given that almost all of the defenders of

objectors assume that everything (including all objects and properties) that exists must do so in time and space and be a part of the causal nexus of the world that can be investigated with the empirical natural sciences. The supervenience challenge captures, in a rigorous form, the skepticism these philosophers have toward views that posit some other kind of entities and properties, such as the *sui generis* normative properties. For them, one reason not to believe in such additional entities and properties is that it would be mysterious how they could be connected to the ordinary natural world in a systematic way as the supervenience challenge argues.

If in this dialectical situation the nonnaturalists' response requires both the discontinuous normative properties and an additional, distinct transcendent realm populated by a set of Platonic universals, the naturalists will reject the view. They will do so already due to the additional metaphysical realm and its entities, which the proposal requires. From the naturalists' perspective, a version of nonnaturalism that can respond to the supervenience challenge but is committed to those things is not any more plausible than a metaphysically more parsimonious version of nonnaturalism that cannot respond to the challenge.[46]

Second, transcendent realism concerning universals has fallen out of favor since Russell's defense of the view due to many well-known powerful objections to it.[47] To see this, consider a case in which an individual has a certain property, say when John has the property of being tall. The main problem for

---

the supervenience challenge have been naturalists, I still do believe that dialectically the responses that do not require accepting any additional nonnaturalist elements beyond the normative properties themselves will be more effective. This is why, even if it is not a knockdown objection to Skarsaune's reliance on transcendent realism, I do think it is an advantage of and motivation for my trope-theoretic proposal below that it relies on a general view of properties that is acceptable for naturalists.

46 It might be worried here that this sets the bar for the nonnaturalist solutions too high: that they must be able to convince the critics of nonnaturalism. My concern about Schroeder's and Skarsaune's proposal is more modest. I merely emphasize that adopting it seems to require a commitment not only to nonnatural properties but also to a separate Platonic realm of abstract entities. Insofar as metaphysical parsimony is a theoretical virtue, such a commitment is a theoretical cost and something that leads to additional objections from the naturalist perspective that go beyond the concerns about the existence of *sui generis* moral properties. One motivation of the view below is that it makes these additional commitments unnecessary.

47 See, e.g., Armstrong, *Nominalism and Realism*, vol. 1, ch. 7; and Edwards, *Properties*, sec. 2.2.3. However, for sympathetic discussions, see Bealer, "Universals and Properties"; MacDonald, *Varieties of Things*; Jubien, *Possibility*; and Van Inwagen, "Relational vs. Constituent Ontologies."

the defenders of transcendent universals is to explain what the relation between John and the tallness universal is here.

The first suggestion is that each particular that shares a given property *participates* (or "partakes") in the universal in question.[48] Yet, it is mysterious how concrete objects that exist in space and time could be "parts" of the universals that are abstract objects in the transcendent realm. Furthermore, because each individual sharing a property would be a different part of a given universal, we would need something further to explain what unifies all these individuals as bearers of the given property.[49] Yet, answering that question was the point of introducing the abstract universals in the first place.

The second suggestion is that the individuals that share a given property *resemble* the relevant universal in some way.[50] Yet, it is difficult to see how this could be, given that individuals are spatio-temporal, concrete, changeable, destructible, and sensible, whereas the universals are nonspatio-temporal, abstract, immutable, indestructible, and insensible.[51] Because of this, some defenders of transcendent universals argue that the relationship between individuals and the abstract universals is primitive—it cannot be explained in any other terms.[52] One important advantage of the trope theory introduced below, however, is that it can explain property instantiation in terms of an ordinary part/whole relation. Insofar as we then have reason to prefer views with fewer theoretical primitives, this is one reason to reject transcendent realism.[53]

5. The Supervenience Challenge and the Trope Theory

This section explores whether we could explain why the relevant action kind first-order properties have their higher-order normative properties necessarily without positing a distinct transcendent realm of abstract universals. Are

48 Russell, "The World of Universals."

49 Armstrong, *Nominalism and Realism*, 1:66.

50 Plato, *The Republic*, 597a.

51 Edwards, *Properties*, 23. The proposal also leads to several third-man-type regresses (Armstrong, *Nominalism and Realism*, vol. 1, ch. 7; Edwards, *Properties*, 23–26).

52 Cook Wilson, *Statement and Inference with Other Philosophical Papers*, sec. 148.

53 See Lewis, *On the Plurality of* Worlds, 154. There is also a concern that transcendent realism threatens to make all higher-order properties to be instantiated necessarily. Yet, some higher-order properties are clearly contingent. The property of redness, for example, has the property of being Jo's favorite color only contingently (see section 7 below; Egan, "Second-Order Predication and the Metaphysics of Properties"; and Cowling, "Intrinsic Properties of Properties"). Transcendent realists, just like the trope theorists discussed below, would thus need to understand all such contingent higher-order properties as mere relations rather than as universals.

there general compelling views of properties that (i) many naturalists already accept, and that (ii) could also provide the missing piece of the puzzle for the nonnaturalist realists?

Trope theory is one of the leading metaphysical theories of properties in analytic ontology.[54] This section introduces it and applies it in the present context. The next section then concludes that, insofar as normative properties are intrinsic properties of action kinds, even in the framework of trope theory the relevant action kind first-order properties have their normative properties in all *possible worlds*, and so the missing element of Schroeder's and Skarsaune's nonnaturalist response can be provided.

According to trope theory, tropes are property instances: the tallness of John is an instance of the property of being tall. This instance is a concrete (it exists in a certain position in time and space), basic particular (it inheres in just one object)—simple, fundamental, and independent.[55] It is a primitive entity called "a trope." Individuals, such as John, are then understood as *bundles* of compresent tropes.[56] As a consequence, a given individual instantiates a prop-

54 Fisher, "*Abstracta* and Abstraction in Trope Theory," 41. For overviews, see Armstrong, *Nominalism and Realism*, vol. 1, ch. 8, and *Universals*, ch. 6; Edwards, *Properties*, ch. 3; Maurin, "Trope Theory"; and Moreland, *Universals*, ch. 3. For defenses, see, e.g., Stout, "The Nature of Universals and Propositions"; Williams, "On the Elements of Being, I–II"; Campbell, "The Metaphysic of Abstract Particulars" and *Abstract Particulars*; Simons, "Particulars in Particular Clothing"; Maurin, *If Tropes* and "Trope Theory and the Bradley Regress"; Ehring, *Tropes*; Schaffer, "The Individuation of Tropes"; and McDaniel, "Tropes and Ordinary Physical Objects." For a more complete list, see Maurin, "Trope Theory," sec. 1.

55 Some trope theorists have argued that the tropes are abstract as at least epistemically we abstract them from individuals (Campbell, "The Metaphysic of Abstract Particulars," 477–78). For reasons not to accept this view, see Simons, "Particulars in Particular Clothing," 557. I follow Maurin in taking tropes to be simple in the sense that they are not constituted of entities belonging to some other categories ("Trope Theory," sec. 2.2; see also Ehring, *Tropes*, 179–80). Furthermore, trope theorists could also think of individuals as bundles of both an individual substance and the compresent tropes (see note 56 below).

56 Campbell, "The Metaphysic of Abstract Particulars," 479, 482–83; Williams, "On the Elements of Being, I–II." Some trope theorists assume that an object consists of a substratum that instantiates the relevant tropes (see Martin, "Substance Substantiated," 7–8). For objections, see Campbell, *Abstract Particulars*, 7; and Daly, "Tropes," 258–59). Compresence is here to be understood as occupying the same point in space and time. This relation can be understood either as an internal or an external relation (see Maurin, "Trope Theory and the Bradley Regress," 321–22, and "Tropes," sec. 3.2). The former alternative seems to make all properties of objects necessary whereas the latter threatens to lead to vicious regresses (see Ehring, *Tropes*, 120–21). Simons suggests that, for this reason, we should think that the tropes that form "the essential kernel or nucleus" of the object are connected by internal relations (and so depend on their existence on the existence of other tropes of the same kind as now in the nucleus), whereas the nonessential property tropes

erty when an instance of that property in part constitutes the individual. The relationship between tropes and individuals is thus the part/whole relation.

Consider then the different instantiations of the same property, such as tallness_1 (in John), tallness_2 (in Paul), and so on. The trope theorists then claim that, as instantiations of the same property, these tropes are *exactly resembling* basic particulars.[57] The "universal" property of tallness can therefore be understood as *the set* of the exactly resembling tropes. Furthermore, to accommodate uninstantiated properties and to avoid the result that the identity of a property depends on how many individuals instantiate it at a given moment, we should think that the relevant trope set that constitutes a given property has as its members not merely all the actual exactly resembling tropes but also all such tropes from all possible worlds.[58] The relationship between the tropes and the corresponding "universals" can thus be understood in terms of standard set membership.

There are, of course, objections to the trope theory that continue to be debated.[59] However, here the theory has several theoretical advantages. First, it continues to be a popular view of properties (see note 54 above). Second, many of its defenders are explicitly metaphysical naturalists, who claim that everything that exists, including all tropes, exists in space and time and is a

---

of the object are related to this core externally ("Particulars in Particular Clothing"). For discussions of whether this solves the problem, see Edwards, *Properties*, 61; and Maurin, "Tropes," sec. 3.2. For other potential solutions to the problem, see Maurin, "Tropes," sec. 3.2.

57 Formally exact resemblance consists of an equivalence relation that is symmetrical, reflexive, and transitive. Here too there is a threat of a regress: this would be the case if two tropes were exactly resembling in virtue of having some more basic exactly resembling tropes (Edwards, *Properties*, 62). Campbell argued that we can avoid this problem by thinking of exact resemblance as an internal relation between tropes determined by their very nature (*Abstract Particulars*, 72). For an objection, see Daly, "Tropes," sec. 3. Another way to avoid the problem is to take exact resemblance as a primitive notion (Campbell, "The Metaphysic of Abstract Particulars," 484; Edwards, *Properties*, 64) or to formulate trope theory in a way that does not rely on exact resemblance (Ehring, *Tropes*, 175).

58 Loux, *Metaphysics*, 83n28. This may seem to commit the view to Lewisian modal realism (see Lewis, *On the Plurality of Worlds*) so that the merely possible tropes can be just as concrete entities as the actual tropes. Yet, we can also think of possible worlds as maximal, consistent descriptions of how the world could be. These descriptions include individual descriptions of particular property instantiations. In this case, the relevant trope set that constitutes a given property has as its members both the actual tropes and the previous descriptions of the merely possible tropes that would exactly resemble them.

59 For many of these debates, see the literature in notes 54–57 above. For an objection to the meaningfulness of trope talk, see Van Inwagen, "Relational vs. Constituent Ontologies," 395. For an overview of these problems and solutions to them, see Maurin, "Trope Theory."

part of the causal nexus of the world to be studied by empirical sciences.[60] Many naturalists, for instance, explicitly rely on tropes to explain the causal powers of objects.[61] In this situation, if nonnaturalists rely on the metaphysical framework provided by the trope theory, this itself cannot be objectionable. Adopting that framework does not bring with it any additional metaphysical commitments, which the naturalists would reject due to their naturalism.[62]

Finally, even if trope theory is metaphysically more parsimonious than transcendent realism, it can still avoid the objections to more austere forms of nominalism.[63] Given that on this view properties are sets of exactly resembling property instances, the view can allow properties to function as the referents of both the singular and predicate terms in sentences such as "Red is a color." In contrast, more austere forms of nominalism need to paraphrase the previous kind of claims in a language that only refers to actual individuals and the sets of which they are members.[64] It is well-known how difficult finding such paraphrases is. Trope theory can avoid these problems as it recognizes that properties exist as a distinct metaphysical category.

60 For naturalism of trope theory, see, e.g., Campbell, *Abstract Particulars*; and Schaffer, "The Individuation of Tropes."

61 Campbell, "The Metaphysic of Abstract Particulars," sec. 3.

62 Here it could be objected that I have not given any reason that would be independent of naturalism to prefer trope theory over transcendent realism. This is because there are objections to both theories (see section 5 and notes 54–57 and 59 above), and both views continue to be defended (see notes 47 and 54). It could thus be objected that dialectically both views are on a par: either equally plausible or implausible. Personally, I do think that trope theory is both more widely accepted and has been developed further to respond to many of the objections to it, but defending the view over transcendent realism is beyond the scope of this article. More modestly, this article can be read as an attempt to show that nonnaturalist realism can be defended against the supervenience challenge not only by relying on transcendent realism but also in the framework of trope theory. This means that the supervenience challenge could have force only if some form of immanent realism about universals were true as the only other alternative, austere nominalism, is problematic for the nonnaturalists for other, more basic reasons (see Jackson, *From Metaphysics to Ethics*, 118–25; and Suikkanen, "Nonnaturalism"). Thus, the more general lesson of this article is that, other than immanent realism about universals, it is difficult to think of any other plausible general account of properties in which the supervenience challenge would have force, which makes the objection less pressing as the objection would require defending immanent realism about properties. Accepting this lesson does not require taking a stand on whether trope theory or transcendent realism is more plausible (and, in fact, one reason that supports these views could be claimed to be that they can be a part of the response to the supervenience challenge).

63 For overviews of these problems, see, e.g., Armstrong, *Universals*, ch. 2; Loux, *Metaphysics*, 52–62; and Edwards, *Properties*, ch. 4.

64 Pap, "Nominalism, Empiricism, and Universals: 1"; Jackson, "Statements about Universals."

Let us then apply trope theory in the present context.[65] Take Ben's and Charlie's identical utterances.[66] Let us assume that these utterance-tokens belong to the same action kind of deliberately insulting utterances (here, we could also choose a more or less fine-grained action kind). If we understand these action tokens as bundles of tropes, one of the tropes that constitutes Ben's utterance is the trope T_1 of instantiating the property of belonging to the previous kind, and one of the tropes that constitutes Charlie's utterance is T_2, where T_1 and T_2 are exactly resembling tropes. The set of all the tropes both in the actual and other worlds that exactly resemble those two tropes is then the first-order property of being a deliberately insulting utterance ($S_{\text{diu}} = \{T_1, T_2, \ldots, T_n\}$).

Following Schroeder's and Skarsaune's suggestion, the previous first-order property, S_{diu}, is the primary bearer of the normative property of being wrong. Translated to trope theory, this is the claim that the set that has all the "being an insulting utterance" tropes as its members itself instantiates a further property of wrongness. In other words, the trope R_1 of instantiating wrongness is one of the tropes that is compresent with the set S_{diu}—the first-order property of being an insulting utterance.[67]

65 Shafer-Landau responded to the supervenience challenge by relying on the idea that normative properties are realized by descriptive properties (*Moral Realism*, 77). Ridge translated this view to the language of trope theory ("Anti-Reductionism and Supervenience," 341–42). According to the resulting view, every normative trope is constituted by a cluster of descriptive tropes even if the normative types are not identical with the descriptive types. Ridge argued that this view fails because it will have to assume the kind of necessary connections between distinct entities that are problematic in the first place ("Anti-Reductionism and Supervenience," 343).

66 See section 3 above.

67 Here the nonnaturalist cannot claim, as many trope theorists would (see McKitrick, "Real Potential," sec. 1.1.1), that the higher-order property of wrongness is the set of the different first-order sets of being certain kinds of an action as then the proposal would collapse into naturalism. This is why the additional wrongness trope is needed at the second-order level (though some trope theorists are skeptical about such higher-order tropes—see, e.g., Heil, *From an Ontological Point of View*, 119). Furthermore, the trope theorists who rely on higher-order tropes to give an account of the higher-order properties have to adopt a level-specific account of how the higher-order tropes constitute higher-order properties via set-membership. This entails that, even if (i) the relevant higher-order trope is compresent with the first-order tropes that belong to the set of exactly resembling tropes that constitute the first-order property (so as to make sense of the relevant instantiation relation) and (ii) compresent first-order tropes generally bundle together to form objects, those higher-order tropes do not bundle together with the other compresent tropes of the lower level to become members of the set of the first-order tropes that constitutes the first-order property. For an objection to trope theory concerning higher-order properties, see Jones, "Nominalist Realism," and for a defense of a higher-order trope theory against this objection, Skiba, "Higher-Order Metaphysics and the Tropes versus Universals Debate."

Two further things need to be noted about this application of trope theory. First, there are also other action kinds that have the property of being wrong such as the actions of shoplifting for fun. Each of these actions, both actual and possible, is in part constituted by a trope of instantiating that very action kind. Call these tropes $P_1, P_2, \ldots, P_n$. The set of these tropes, $S_{\text{sfs}} = \{P_1, P_2, \ldots, P_n\}$, is then the property of being an action of shoplifting for fun. This property, too, instantiates wrongness, and so it would have the trope R_2 as one of the relevant compresent tropes.

This means that the property of wrongness would be the set of all the instances of wrongness (i.e., wrongness tropes) that all the different action kinds that are wrong have. It would be the set $S_{\text{wrong}} = \{R_1, R_2, \ldots, R_n\}$. The previous metaphysical picture allows us also to formulate the metaethical disagreement between naturalists and nonnaturalists. The naturalists will claim that the relevant instances of wrongness (tropes $R_1, R_2, \ldots, R_n$) have general properties, such as belonging to the subject matter of sciences, being *a posteriori* detectable, having causal powers, and so on just like all the other ordinary natural properties. In contrast, the nonnaturalists will argue that the wrongness tropes do not instantiate those properties but rather their opposites, which makes the property of wrongness, i.e., the set S_{wrong}, a different kind of a property.

Applying the trope theory to action kinds and their normative properties then provides a new framework for formulating Schroeder's and Skarsaune's nonnaturalist response to the supervenience challenge. On this view, the first-order property of being an action of a certain kind is one transworld entity spread across all possible worlds. It is the set of all the "being that specific kind of an action" tropes that can be found from different possible worlds where that kind of action is done. As the members of that set—the relevant action kind tropes—are spread across all possible worlds, the resulting set that constitutes the property of being that kind of an action, too, is a single entity spread across all worlds. Now, either this first-order property (i.e., the action kind as the set of the relevant tropes) instantiates a given normative property, or it does not. If it does, there is only one case to consider: the one set spread across all *possible worlds*. This means that it cannot be that a given action kind, say being a deliberately insulting utterance, only in some possible worlds has the property of being wrong. The fact that, if the kind has that normative property, it has it in all worlds hence follows from the account of the nature of the relevant first-order action kind property—from it being one set of property instances spread across all worlds.

This feature of set-theoretic accounts of properties according to which the members of those properties are spread across all worlds is well-known. Lewis, for example, thought that, instead of property instantiations, different

properties are sets of both actual and possible individuals. He was aware that, because properties are as a result transworld entities—literally identical across all worlds—properties have their higher-order properties necessarily.[68] As Lewis puts it, "[a] universal can safely be part of many worlds because it hasn't any accidental intrinsics."[69]

The set-theoretic trope theory thus entails that if a first-order property has a certain second-order property, it has that property necessarily. Together with Schroeder's and Skarsaune's proposal, this enables the nonnaturalists to explain supervenience. Claims about the normative properties of action tokens are mixed claims according to which (i) the token belongs to a certain kind and (ii) that kind has a certain normative property. We know that action tokens that share all the same nonnormative base properties must belong to the same action kinds, and the previous account of properties entails that if an action kind has a normative property, it has it necessarily.

6. Two Objections, Responses, an Amendment, and a Conclusion

There are, however, two important objections to the previous proposal, and the response to the second one especially has an interesting consequence for how nonnaturalist realism should be formulated. The nonnaturalists will have to take normative properties to be intrinsic properties of action kinds.[70]

6.1. An Alternative Account

The first objection is based on an alternative trope-theoretic account of normative properties.[71] On this view, the primary bearers of the normative tropes are particular first-order descriptive tropes. Ben's utterance, for example, would, according to this view, have as its part a descriptive trope of deliberately insulting someone, which would then bear the second-order normative trope of being wrong. The generalization expressed by "deliberately insulting someone is wrong" would then be true because all actual and possible tropes

68 Lewis, *On the Plurality of Worlds*, 205; Egan, "Second-Order Predication and the Metaphysics of Properties," 49–50.

69 Lewis, *On the Plurality of* Worlds, 205n6. Lewis did not think that this was a problem because he did not think that there were any good examples of accidental intrinsic higher-order properties, whereas accidental relational higher-order properties can be dealt with in a way discussed below.

70 Schroeder, "The Price of Supervenience," 141–42.

71 I thank an anonymous referee of the *Journal of Ethics and Social Philosophy* for raising this concern. In outlining the objection, the first four paragraphs of this section draw heavily from his or her comments.

of deliberately insulting someone would bear a second-order trope of being wrong.

It could furthermore be argued that there are three good reasons to accept this trope-theoretic view of normative properties rather than the one outlined in section 5 above. First, it arguably better fits the idea that the fundamental wrong-making features of actions are their descriptive qualities (i.e., first-order tropes) rather than any facts about to which sets they belong. The ground of Ben's having acted wrongly seems to be his having insulted someone rather than him doing an action of a certain category.

Second, the view seems supported by what many trope theorists claim about the relata of causal relations.[72] According to them, the first-order tropes themselves are the basic relata in causal relations directly and not in virtue of what kind of tropes they are, and causal generalizations are derivative of the facts about these relations. If we then agree that tropes themselves (rather than sets thereof) do the primary causal work throughout the universe, it seems tempting to suppose also that the first-order tropes do the primary wrong-making as well (notwithstanding the greater modal strength of the latter kind of relation).

Finally, it could be argued that the idea that a set could be a bearer of wrongness in anything other than a derivative sense is a category mistake. After all, it is awkward to say that the property (which is a set of tropes) of being a deliberative insult is wrong, whereas it is not awkward to say that deliberate insults are the kind of actions that are wrong. It could be suggested that we should thus prefer this alternative trope-theoretic view to my proposal, and so that proposal cannot be used in a compelling nonnaturalist response to the supervenience challenge.

There are several things to be said in response to this objection. The first is that the previous proposal can explain neither strong nor weak supervenience. It cannot explain strong supervenience as each of the descriptive tropes of different actions are world-bound phenomena. It also cannot explain weak supervenience because each trope is numerically distinct from the other members in its resemblance class, and the nonnatural normative properties (and tropes) are distinct existences from the descriptive properties (and tropes). In this case, there would be no explanation of why it could not be that one trope of being a deliberative insult bears the wrongness trope while another would not. It could then be argued that philosophical hypotheses are supported by their problem-solving and explanatory power. That my proposal can help the nonnaturalist realist to explain how normative properties supervene on the base properties is itself at least some reason to prefer that proposal over the alternative trope-theoretic proposal that cannot do so.

72 Campbell, *Abstract Particulars*, 113; Ehring, *Causation and Persistence*, ch. 3.

Second, the proposal is also supported by section 2's independent arguments (and the arguments provided by Schroeder and Skarsaune—see note 25 above) for the conclusion that the primary bearers of normative properties are action kinds. These general arguments are neutral about how we should understand properties, but if we want to capture how they instruct us to understand the bearers of normative properties, then within the trope theory, the only consistent option is to think that the bearers of normative properties are to be understood as sets of tropes.

Third, the proposal outlined in section 5, too, is compatible with the idea that the fundamental wrong-makers are their descriptive qualities, the first-order tropes, rather than any facts about which set they belong to. This is because, insofar as we understand wrong-making in terms of metaphysical grounding, it, too, will be a transitive relation.[73] Thus, if a particular action belongs to the kind of deliberative insults in virtue of its first-order descriptive properties, and belonging to that kind makes the action wrong, then by transitivity, the fundamental wrong-makers of the action will be its first-order descriptive qualities.

In responding to the disanalogy of the causal relata objection, there are two options. First, it is possible to defend the idea that the causal and normative realms are genuinely different in structure. This is because, even if the arguments in section 2 give us good reasons to think that the primary bearers of normative properties are action kinds, there are no corresponding arguments with respect to causal relata. There we have better reasons to think that the causal relata are basic first-order tropes unmediated by any set membership.[74] And so, given how these arguments point in different directions, we should recognize differences where they exist.

The second alternative is to argue that the two realms are, in fact, more analogous than the objection suggests. Consider Frank Jackson and Philip Pettit's example of a glass cracking because it contains hot water.[75] In this case, we can think of the temperature of the water as a higher-order property that is realized in this case by certain water molecules having a certain momentum. Here, even if the momentum of these molecules, rather than the higher-order property, causes the glass to crack, the temperature property is still causally relevant as it can be cited in a good causal explanation. This is because the presence of the temperature property ensures that "there would be some property there to exercise the efficacy required."[76]

73 Fine, "Guide to Ground."

74 Ehring, *Causation and Persistence*, ch. 3.

75 Jackson and Pettit, "Program Explanation," 110.

76 Jackson and Pettit, "Program Explanation," 114.

The proposal made in section 5 can be formulated to be analogous to this model of higher-order program explanations in nature. We can think of the property of belonging to an action kind to be like the temperature property in the previous case. In the same way as the action kind is the primary bearer of the normative property (say, wrongness), the temperature property can be thought of as the bearer of the property of explaining why the glass cracked. In addition, just as the temperature property does not itself cause the glass to crack but rather ensures that there is some more basic first-order momentum property that does so, similarly it could be claimed that the belonging to the action kind property itself does not make the action wrong, but rather it merely ensures that the action has some more basic, first-order descriptive property that does so. We can make this claim if we think that the action kind bears its normative property in virtue of the first-order properties of all its instances. With this picture, it could be argued that the causal and normative realms turn out to be analogous in a way that blocks the previous objection.

Finally, with respect to the category mistake charge, it is important to keep in mind what the proposal is a proposal of. It is true that, in everyday life, claiming that it is wrong to deliberately insult someone sounds natural, whereas any claims about the property of being a deliberative insult (or a set of tropes) being wrong sounds just confused. But, to some degree, this reaction is to be expected. The whole point of the proposal is to make sense of the former type of ordinary claims by making explicit their truth conditions. This is done in two stages. In the first stage, the ordinary claim is analyzed in terms of the action token belonging to an action kind and the kind instantiating the relevant normative property. Then, in the second stage, we attempt to provide a trope-theoretic metaphysical theory of what it is for the action kind to be instantiating that normative property in a way that can also explain supervenience. It is not surprising that at this point we may end up saying things that do not sound right in the ordinary language, but this happens in metaphysics relatively often anyway.

To see this, consider ordinary modal claims such as "Tim can open the door." According to Lewis, the truth conditions of this claim are provided by whether Tim has a counterpart, a person very much like him but not numerically identical to him, in a different possible world who opens a similar door there.[77] At this point, it could be objected that this analysis commits a category mistake as the original claim is about Tim and what he can do in this world, whereas the latter claim is about what a different person altogether can do somewhere else.[78] But, here too, we should expect that the account of the truth conditions

77 Lewis, *Counterfactuals*, 39–40.

78 Kripke, *Naming and Necessity*, 45n13.

of the ordinary claims themselves might not be intuitive, and yet whether we should accept the view should depend more on the explanatory power of the account overall.[79]

6.2. Intrinsic and Relational Higher-Order Properties

The second objection begins from the thought that intuitively there are contingent higher-order properties, which first-order properties have in some worlds but not in others.[80] Yet, we cannot make sense of such properties in the previous framework. One example is the property of being somebody's favorite color.[81] In our world, greenness has this property, but there are worlds where green is not anybody's favorite color. According to the previous proposal, the property of being green is the set of all actual and possible instantiations of greenness. This set is one transworld entity—identical in every world. To say that this one entity would both have and not have the property of being somebody's favorite color would be a contradiction.

This means that, according to the previous framework, even this higher-order property could not be contingent, and yet clearly it is. Furthermore, if we respond to this objection by amending the trope-theoretic framework in a way that it will be able to accommodate contingent higher-order properties, the original concern returns. The opponents of nonnaturalism can argue that the set of tropes that constitutes a certain action kind will be a first-order property that could well have its higher-order normative properties contingently, and so Schroeder's and Skarsaune's response would fail in the way explained in section 3.

Trope theorists have one strategy for making room for contingent higher-order properties. It begins from recognizing that there are both instances of monadic properties and relations, i.e., instances of relational properties. The former instances give rise to monadic properties (sets of monadic tropes), whereas the latter to relational properties (sets of relational tropes). We can then think of the relational tropes as mere relations in disguise—they are roughly the relations that in some way connect entities that are not dependent on one another.[82] More precisely, the existence of relational tropes depends on the very tropes they relate, whereas the existence of monadic tropes does not depend on the existence of some specific tropes, be they relational or not.[83]

79 Lewis, *Counterfactuals*, ch. 4.

80 Cowling, "Intrinsic Properties of Properties," 244.

81 Egan, "Second-Order Predication and the Metaphysics of Properties," 49.

82 Because of this stipulation, relational tropes cannot connect the tropes of a bundle that constitutes a certain individual.

83 See, e.g., McDaniel, "Tropes and Ordinary Physical Objects," 271–72.

Consider then the previous example.[84] Here we are understanding the property of greenness as a transworld entity, as a set of all actual and possible instances of greenness. This property has two relational properties: being somebody's favorite color in the actual world @ and not being anybody's favorite in world w_n. We can then understand these relational properties as relational tropes (which are really relations in disguise). The one transworld entity of greenness has the relational trope of being suitably related to the favoring attitudes of different individuals in @ and the relational trope of not being suitably related to anyone's color preferences in w_n. What we then mean when we say that green is somebody's favorite color when we take this to be a contingent claim is that greenness is suitably related to some people's color attitudes in our world but not in others.[85]

Yet, it could be argued that, instead of consisting of relations to other things, monadic tropes are intrinsic to an individual (that itself is a bundle of tropes). According to one attractive version of this type of a trope theory, these intrinsic tropes are either (1) one of the mutually dependent tropes that constitute the "nucleus" of the individual or (2) one of the tropes that constitute the "halo" of the individual that the individual has in virtue of only the previous tropes that make up the nucleus (the tropes of the nucleus can at most depend on the existence of the same kind of tropes as the ones in the halo of the bundle but not on those specific tropes).[86] The existence of these types of individuals constituting monadic tropes then does not depend on the existence of the tropes that constitute any other individual.

Return then to the action kinds as first-order properties of action tokens and the normative properties as their higher-order properties. Consider the property of being a deliberately insulting utterance. Within the framework provided, this property is the set of both actual and possible tropes of being that kind of an action. If we take normative properties, such as wrongness, to be relational properties of the previous type of a set, then whether the action kind instantiates the property of wrongness is contingent. For example, if we thought that the wrongness of uttering insults depends on which moral code is accepted in

84 Here I follow Egan, "Second-Order Predication and the Metaphysics of Properties," 50–51; and Lewis, *On the Plurality of Worlds*, 201.

85 Egan objects that this way of understanding the semantic content of the claim will still make the content come out objectionably as necessarily true ("Second-Order Predication and the Metaphysics of Properties," 50–51). He argues that if we think of properties as functions from worlds to extensions, this problem is avoided ("Second-Order Predication and the Metaphysics of Properties," sec. 3). The view below can be translated to this language if we think of the relevant extensions as the transworld sets of tropes.

86 See, e.g., Simons, "Particulars in Particular Clothing." Here in 2 I rely on a hyperintensional view of intrinsicality based on the "in virtue of" relations, as defended by Bader ("Towards a Hyperintensional Theory of Intrinsicality").

a world, the provided framework would have room for the way in which the wrongness of insulting would be a contingent higher-order property. The transworld set of actual and possible instances of being an insulting utterance would in this case instantiate one relation to the conventional morality of the actual world (the relation of being forbidden by) and a different relation to the conventional moralities of some other worlds (the relation of being authorized by).

Of course, the nonnaturalist realists do not accept that account as they think that whether it is wrong to make insulting utterances is a stance-independent fact. On their view, for something to instantiate a normative property is not a question of being related in some way to the conventional morality of a community. Nonnaturalist realists thus think that whether a certain action kind instantiates wrongness depends only on the qualities of that action kind (the first-order tropes that are the constitutive members of the action kind set) and whether those qualities are wrong-makers, and not on anything else. They thus think that normative properties, such as wrongness, are intrinsic monadic properties of action kinds.

Yet, fortunately for the nonnaturalist realists, according to the outlined trope-theoretic framework, first-order properties have their intrinsic second-order properties necessarily simply in virtue of the general nature of properties (i.e., in virtue of the nature of the first-order exactly resembling tropes that are the members of the set that is the given first-order property). Hence, insofar as the nonnaturalists take normative properties to be intrinsic properties of action kinds, they have a response to the supervenience challenge. They can argue that the first-order property of belonging to a certain kind of actions is a single transworld entity, a set of the first-order tropes, that instantiates a given intrinsic higher-order intrinsic normative property (in virtue of the second-order normative property trope being compresent with a relevant set of the first-order tropes). As a result of this trope-theoretic framework, the action kind will have the normative property necessarily (across all *possible worlds*), and so it cannot be the case that different action tokens that have the same base properties (and thus ones that belong to the same action kinds) could have different normative properties.

University of Birmingham
j.v.suikkanen@bham.ac.uk

REFERENCES

Anscombe, G. E. M. *Intention*. Oxford: Basil Blackwell, 1957.
Armstrong, David M. *Nominalism and Realism: Universals and Scientific Realism*.

Vol. 1. Cambridge: Cambridge University Press, 1978.
———. *Universals: An Opinionated Introduction*. New York: Westview, 1989.
Bader, Ralf M. "Towards a Hyperintensional Theory of Intrinsicality." *Journal of Philosophy* 110, no. 10 (October 2013): 525–63.
Bealer, George. "Universals and Properties." In *Contemporary Readings in the Foundations of Metaphysics*, edited by Stephen Laurence and Cynthia MacDonald, 131–47. Oxford: Blackwell, 1998.
Blackburn, Simon. *Essays in Quasi-Realism*. Oxford: Oxford University Press, 1993.
———. *Ruling Passions: A Theory of Practical Reasoning*. Oxford: Oxford University Press, 1998.
Campbell, Keith. *Abstract Particulars*. Oxford: Blackwell, 1990.
———. "The Metaphysic of Abstract Particulars." *Midwest Studies in Philosophy* 6, no. 1 (September 1981): 477–88.
Cook Wilson, John. *Statement and Inference with Other Philosophical Papers*. Oxford: Clarendon, 1926.
Copp, David. "Why Naturalism?" *Ethical Theory and Moral Practice* 6, no. 2 (June 2003): 179–200.
Cowling, Sam. "Intrinsic Properties of Properties." *Philosophical Quarterly* 67, no. 267 (April 2017): 241–62.
Cudworth, Ralph. *A Treatise Concerning Eternal and Immutable Morality*. Edited by Sarah Hutton. Cambridge: Cambridge University Press, 1996.
Cuneo, Terence. *The Normative Web: An Argument for Moral Realism*. New York: Oxford University Press, 2010.
Daly, Chris. "Tropes." *Proceedings of the Aristotelian Society* 94, no. 1 (June 1994): 253–62.
Dancy, Jonathan. *Practical Shape: A Theory of Practical Reasoning*. Oxford: Oxford University Press, 2018.
Dreier, James. "Another World: The Metaethics and Metametaethics of Reasons Fundamentalism." In *Passions and Projections: Themes from the Philosophy of Simon Blackburn*, edited by Robert N. Johnson and Michael Smith, 154–71. Oxford: Oxford University Press, 2015.
———. "Explaining the Quasi-Real." In *Oxford Studies in Metaethics*, vol. 10, edited by Russ Shafer-Landau, 273–98. Oxford: Oxford University Press, 2015.
———. "Is There a Supervenience Problem for Robust Moral Realism?" *Philosophical Studies* 176, no. 6 (June 2019): 1391–408.
———. "The Supervenience Argument against Moral Realism." *The Southern Journal of Philosophy* 30, no. 3 (Fall 1992): 13–38.
Dunaway, Billy. "Epistemological Motivations for Anti-Realism." *Philosophical*

Studies 175, no. 11 (November 2018): 2763–89.
Edwards, Douglas. *Properties*. Cambridge: Polity, 2014.
Egan, Andy. "Second-Order Predication and the Metaphysics of Properties." *Australasian Journal of Philosophy* 82, no. 1 (March 2004): 48–66.
Ehring, Douglas. *Causation and Persistence: A Theory of Causation*. Oxford: Oxford University Press, 1997.
———. *Tropes: Properties, Objects, and Mental Causation*. Oxford: Oxford University Press, 2011.
Enoch, David. *Taking Morality Seriously: A Defence of Robust Realism*. Oxford: Oxford University Press, 2011.
Fine, Kit. "Guide to Ground." In *Metaphysical Grounding*, edited by Fabrice Correia and Benjamin Schnieder, 37–80. Cambridge: Cambridge University Press. 2012.
———. "Necessity and Nonexistence." In *Modality and Tense: Philosophical Papers*, 321–54. Oxford: Oxford University Press, 2005.
———. "Varieties of Necessity." In *Conceivability and Necessity*, edited by Tamar Szabó Gendler and John Hawthorne, 253–81. Oxford: Oxford University Press, 2002.
Fisher, A. R. J. "*Abstracta* and Abstraction in Trope Theory." *Philosophical Papers* 49, no. 1 (March 2020): 41–67.
FitzPatrick, William Joseph. "Robust Ethical Realism, Nonnaturalism, and Normativity." In *Oxford Studies in Metaethics*, vol. 3, edited by Russ Shafer-Landau, 159–205. Oxford: Oxford University Press, 2008.
Gibbard, Allan. *Reconciling Our Aims: In Search of Bases for Ethics*. Oxford: Oxford University Press, 2008.
———. *Thinking How to Live*. Cambridge, MA: Harvard University Press, 2003.
———. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, MA: Harvard University Press, 1990.
Goldman, Alvin I. *A Theory of Human Action*. Englewood Cliffs, NJ: Prentice Hall, 1970.
Hare, R. M. *The Language of Morals*. Oxford: Oxford University Press, 1952.
———. "The Simple Believer." In *Essays on Religion and Education*, 1–39. Oxford: Clarendon Press, 1992.
Heil, John. *From an Ontological Point of View*. Oxford: Oxford University Press, 2003.
Huemer, Michael. *Ethical Intuitionism*. New York: Palgrave MacMillan, 2005.
Jackson, Frank. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press, 1998.
———. "Statements about Universals." *Mind* 86, no. 343 (July 1977): 427–29.
Jackson, Frank and Philip Pettit. "Program Explanation: A General Perspective."

Analysis 50, no. 2 (March 1990): 107–17.
Jones, Nicholas K. "Nominalist Realism." *Noûs* 52, no. 4 (December 2018): 808–35.
Jubien, Michael. *Possibility*. Oxford: Oxford University Press, 2009.
Kramer, Matthew H. "Supervenience as an Ethical Phenomenon." *American Journal of Jurisprudence* 50, no. 1 (June 2005): 173–224.
Kripke, Saul A. *Naming and Necessity.* Oxford: Blackwell, 1981.
Leary, Stephanie. "Nonnaturalism and Normative Necessities." In *Oxford Studies in Metaethics*, vol. 12, edited by Russ Shafer-Landau, 76–105. Oxford: Oxford University Press, 2017.
Lewis, David. *Counterfactuals*. Oxford: Blackwell, 1973.
———. "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61, no. 4 (December 1983): 343–77.
———. *On the Plurality of Worlds.* Oxford: Blackwell, 1986.
Little, Margaret. "Moral Realism II: Nonnaturalism." *Philosophical Books* 35, no. 4 (October 1994): 225–32.
Loux, Michael J. *Metaphysics: A Contemporary Introduction.* 3rd ed. London: Routledge, 2006.
MacDonald, Cynthia. *Varieties of Things: Foundations of Contemporary Metaphysics*. Oxford: Oxford University Press, 2005.
Martin, C. B. "Substance Substantiated." *Australasian Journal of Philosophy* 58, no. 1 (March 1980): 3–10.
Maurin, Anna-Sofia. *If Tropes*. Dordrecht: Kluwer, 2002.
———. "Trope Theory and the Bradley Regress." *Synthese* 175, no. 3 (August 2010): 311–26.
———. "Tropes." *Stanford Encyclopedia of Philosophy* (Summer 2018). https://plato.stanford.edu/archives/sum2018/entries/tropes/.
McDaniel, Kris. "Tropes and Ordinary Physical Objects." *Philosophical Studies* 104, no. 3 (June 2001): 269–90.
McKitrick, Jennifer. "Real Potential." In *Handbook of Potentiality*, edited by Kristina Engelhard and Michael Quante, 229–60. Dordrecht: Springer, 2018.
McPherson, Tristram. "Ethical Nonnaturalism and the Metaphysics of Supervenience." In *Oxford Studies in Metaethics*, vol. 7, edited by Russ Shafer-Landau, 205–34. Oxford: Oxford University Press, 2012.
———. "Supervenience in Ethics." *Stanford Encyclopedia of Philosophy* (Winter 2015). https://plato.stanford.edu/archives/win2015/entries/supervenience-ethics/.
Moore, G. E. *Principia Ethica*. Cambridge: Cambridge University Press, 1903.
Moreland, James Porter. *Universals*. Chesham: Acumen, 2001.
Olson, Jonas. *Moral Error Theory*: *History, Critique, Defence*. Oxford: Oxford

University Press, 2014.
Pap, Arthur. "Nominalism, Empiricism, and Universals: 1." *Philosophical Quarterly* 9, no. 37 (October 1959): 330–40.
Plato. *The Republic*. Edited by G. R. F Ferrari. Translated by Tom Griffin. Cambridge: Cambridge University Press, 2000.
Price, Richard. *A Review of the Principal Questions in Morals*. Edited by D. D. Raphael. Oxford: Oxford University Press, 1948.
Prichard, H. A. "Duty and Ignorance of Fact." In *Moral Writings*, edited by Jim MacAdam, 84–101. Oxford: Oxford University Press, 2002.
Ridge, Michael. "Anti-Reductionism and Supervenience." *Journal of Moral Philosophy* 4, no. 3 (January 2007): 330–48.
Rosen, Gideon. "Metaphysical Relations in Metaethics." In *The Routledge Handbook of Metaethics*, edited by Tristram McPherson and David Plunkett, 251–69. New York: Routledge, 2017.
Russell, Bertrand. "The World of Universals." In *The Problems of Philosophy*, 142–57. London: Williams and Norgate, 1912.
Scanlon, T. M. *Being Realistic about Reasons*. Oxford: Oxford University Press, 2014.
Schaffer, Jonathan. "The Individuation of Tropes." *Australasian Journal of Philosophy* 79, no. 2 (June 2001): 247–57.
Schmitt, Johannes, and Mark Schroeder. "Supervenience Arguments under Relaxed Assumptions." *Philosophical Studies* 155, no. 1 (August 2011): 133–60.
Schroeder, Mark. "The Price of Supervenience." In *Explaining the Reasons We Share*, 124–44. Oxford: Oxford University Press, 2014.
Shafer-Landau, Russ. *Moral Realism: A Defence*. Oxford: Oxford University Press, 2003.
Shoemaker, Sydney. "Causality and Properties." In *Time and Cause: Essays Presented to Richard Taylor,* edited by Peter van Inwagen, 109–35. Dordrecht: Reidel, 1980.
Simons, Peter. "Particulars in Particular Clothing: Three Trope Theories of Substance." *Philosophy and Phenomenological Research* 54, no. 3 (September 1994): 553–75.
Skarsaune, Knut Olav. "How to Be a Moral Platonist." In *Oxford Studies in Metaethics*, vol. 10, edited by Russ Shafer-Landau, 245–72. Oxford: Oxford University Press, 2015.
Skiba, Lukas. "Higher-Order Metaphysics and the Tropes versus Universals Dispute." *Philosophical Studies* 178, no. 9 (September 2021): 2805–27.
Stocker, Michael. "Duty and Supererogation." *American Philosophical Quarterly* Monograph no. 1 (1968): 53–63.
Stout, G. F. "The Nature of Universals and Propositions." *Proceedings of the*

British Academy 10 (1921): 157–72.
Stratton-Lake, Philip, and Brad Hooker. "Scanlon versus Moore on Goodness." In *Metaethics after Moore*, edited by Mark Timmons, 149–68. Oxford: Oxford University Press, 2006.
Sturgeon, Nicholas. "Doubts about the Supervenience of the Evaluative." In *Oxford Studies in Metaethics*, vol. 4, edited by Russ Shafer-Landau, 53–92. Oxford: Oxford University Press, 2009.
Suikkanen, Jussi. "Nonnaturalism: The Jackson Challenge." In *Oxford Studies in Metaethics*, vol. 5, edited by Russ Shafer-Landau, 87–110. Oxford: Oxford University Press, 2010.
Swoyer, Chris. "The Nature of Natural Laws." *Australasian Journal of Philosophy* 60, no. 3 (September 1982): 203–23.
Toppinen, Teemu. "Nonnaturalism Gone Quasi: Explaining the Necessary Connections between the Natural and the Normative." In *Oxford Studies in Metaethics*, vol. 13, edited by Russ Shafer-Landau, 25–47. Oxford: Oxford University Press, 2018.
Vallentyne, Peter. "The Nomic Role Account of Carving Reality at the Joints." *Synthese* 115, no. 2 (May 1998): 171–98.
Van Cleve, James. "Brute Necessity." *Philosophy Compass* 13, no. 9 (September 2018): 1–43.
Van Inwagen, Peter. "Relational vs. Constituent Ontologies." *Philosophical Perspectives* 25, no. 1 (December 2011): 389–405.
Väyrynen, Pekka. "The Supervenience Challenge to Nonnaturalism." In *The Routledge Handbook of Metaethics*, edited by Tristram McPherson, 170–84. New York: Routledge, 2018.
Wedgwood, Ralph. "The Price of Non-Reductive Moral Realism." *Ethical Theory and Moral Practice* 2, no. 3 (September 1999): 199–215.
Wielenberg, Erik J. *Robust Ethics: The Metaphysics and Epistemology of Godless Normative Realism*. Oxford: Oxford University Press, 2014.
Williams, Donald C. "On the Elements of Being: I." *Review of Metaphysics* 7, no. 1 (September 1953): 3–18.
———. "On the Elements of Being: II." *Review of Metaphysics* 7, no. 2 (December 1953): 171–92.

Journal of Ethics and Social Philosophy
VOL. 26, NO. 3 · FEBRUARY 2024
https://doi.org/10.26556/jesp.v26i3.3086

PRIVILEGED CITIZENS AND THE RIGHT TO RIOT

A REPLY TO PASTERNAK

Thomas Carnes

Sometimes, "political rioting … can be justified in democracies under circumstances that are not far from the reality of many states in the world."[1] This is the conclusion Avia Pasternak reaches in her article "Political Rioting: A Moral Assessment." I wholeheartedly agree with this conclusion. But in reaching this conclusion, Pasternak restricts who, on her account, may permissibly participate in such justified riots. Specifically, Pasternak insists that only genuinely oppressed citizens may permissibly riot. Due to a difference in political circumstances, which will be discussed below, privileged citizens may not permissibly riot on Pasternak's account.

This discussion note argues that such a constraint should be eliminated from an account of permissible riots. I argue, specifically, that Pasternak's account is able to accommodate the permissibility of privileged citizens rioting and that doing so improves her account on its own terms. I first lay out the definition and understanding of political rioting that Pasternak uses before discussing the conditions her account imposes on rioters. Understanding what she takes to constitute a political riot and how it differs from other forms of violence (political or otherwise) will be important to the rest of my argument. I then argue why privileged citizens can be justified in rioting alongside oppressed citizens.

1. WHAT IS POLITICAL RIOTING?

Pasternak defines a political riot as "a public disorder in which a large group of actors, acting spontaneously and without formal organization, engages in acts of lawlessness and open confrontation with law enforcement agencies."[2]

1 Pasternak, "Political Rioting," 418.

2 Pasternak, "Political Rioting," 388. This is admittedly open ended. I do not have space here to more fully flesh out the notion of rioting, and I do not think much in my argument turns on any such considerations. While it would surely be useful and important to examine

Riots are typically a response of an oppressed group to the shared experience of "subjective deprivation, social exclusion, political powerlessness, and moral outrage."[3] Pasternak's account is importantly limited to rioters responding to "severe and pervasive social injustices," such as that reflected by the persistence of the American urban ghetto.[4] Riots typically involve violence or harm carried out specifically with the aim of bringing about political changes that will "eradicate, or in the least ameliorate, the substantive violations of justice" rioters take the state to be responsible for.[5] This feature of political rioting distinguishes it from "maddened" or "senseless or opportunistic" violence, which is often how political rioting is characterized.[6] It is this characterization that leads to many commentators offering the kinds of blanket condemnation of rioters that Pasternak's account of permissible rioting intends to refute.

The key feature here is that permissible political rioting seeks specifically to eliminate or ameliorate the gross injustices to which the rioting Pasternak and I have in mind respond.[7] Pasternak maintains that for rioters to be justified in rioting, they must "remain fundamentally committed to the realization of the democratic ideal."[8] This combination of the presence of severe injustice and

---

different forms of rioting in a more systematic manner, I take any such examination to be beyond the scope of this paper. My focus here is not on the merits or demerits of particular actions rioters might take but rather the antecedent question (in my view, anyway) of which citizens satisfy the conditions that must be met to justify the resort to any sort of "public disorder … lawlessness and open confrontation" in response to "severe and pervasive social injustices." As a result, I feel comfortable leaving aside concerns about the open-endedness of Pasternak's notion of rioting. I would like to thank an anonymous reviewer for pressing this point.

3 Waddington, "The Madness of the Mob?," 681.

4 Pasternak, "Political Rioting," 389.

5 Pasternak, "Political Rioting," 392.

6 Pasternak, "Political Rioting," 391, 389.

7 It is important to note that this is an objective criterion and not one the satisfaction of which turns on whether a rioter considers himself oppressed or otherwise subject to "gross injustice." I, therefore, do not take it to be the case that much rides (here, at least) on the distinction between "oppressed" and "privileged." Given the brutal history in the United States, for example, of decidedly privileged citizens rioting in response to objectively misguided and immoral senses of injustice, there is a need for a clear distinction between oppressed and privileged citizens for any comprehensive account of rioting, especially one that seeks to defend at least some participation of privileged citizens. However, I regrettably lack the space for this here, leaving such hard questions for later consideration. This paper assumes the presence of clear and undeniably gross injustice, and I think starting with such "easy cases" will, in fact, help us to address these hard questions at the level of specificity warranted by such an important issue once the contours of a plausible account of permissible rioting have been developed.

8 Pasternak, "Political Rioting," 395.

a commitment to the democratic ideal—albeit one that manifests in destructive acts of public defiance—is necessary for political rioting to be justifiable. While Pasternak does not articulate the point in this way, I take it that these two conditions jointly ensure that political rioters have what we might call "just cause" to engage in rioting. Riots with a just cause are the only ones I consider in this paper.

On Pasternak's view, for a given case of rioting to be justified, it must meet two further conditions—a success condition and a necessity condition: the rioting must have a reasonable chance of successfully achieving its just aims, and it must be necessary in order to achieve its just aims. In the next section, I lay out the success and necessity conditions and argue that the necessity condition is too narrow insofar as it implies that only citizens who suffer the injustices that give rise to rioting's just cause may permissibly riot. I argue that other citizens can be justified in participating in a permissible riot and that expanding the scope of Pasternak's view in this way will actually increase the chances of success.

2. THE SUCCESS AND NECESSITY CONDITIONS

Pasternak's account of permissible rioting adopts a defensive violence framework, acknowledging that while defensive violence can be justified, it must meet certain conditions.[9] The first condition Pasternak lays out is that political rioting must have a "reasonable prospect to avert, or in the least ameliorate the attack that triggered it."[10] In arguing that political rioting in response to severe injustice can possibly meet this condition, Pasternak relies on empirical evidence to suggest riots can, in fact, play an important role in bringing about positive policy changes that constitute substantive amelioration (if not elimination) of the injustices to which the rioting responds. Given the difficulty of creating genuine social change, even just ameliorating the injustices can constitute a significant victory in the fight for justice. A compelling example is the

9 Pasternak's account, following accounts of defensive harm, rests on three conditions: success, necessity, and proportionality. In what follows, I set aside proportionality because my argument regarding the permissibility of privileged citizens rioting does not turn on considerations of proportionality like it does on considerations of success and necessity. If resorting to rioting as a response to gross injustice, as opposed to other kinds of response, is itself disproportionate to the injustice to which the rioting responds, even if it has a reasonable chance of success and is necessary to eliminate or ameliorate the injustice to which it responds, then neither oppressed nor privileged citizens will be permitted to riot. If rioting is proportionate, then whether privileged citizens may riot will turn on other considerations.

10 Pasternak, "Political Rioting," 398.

US race riots of the 1960s that led to the Kerner Report, which "had a substantive impact on federal aid programs to inner city populations."[11] It is not only that these policy changes were substantive changes, but they were changes that had a real (though certainly incomplete) ameliorative effect on some of the injustices that the 1960s race riots were responding to. This case demonstrates that rioting can indeed achieve at least some of its aims, bolstering Pasternak's argument that rioting can indeed be justifiable.

In addition to having a reasonable chance of success, political rioting must also be necessary. To meet this necessity condition, it must be the case that there are no other less violent ways to bring about the policy changes that rioting seeks to bring about.[12] An obvious objection to the justifiability of rioting insists that, at least in the democratic societies Pasternak's account focuses on, the very nature of democracy provides multiple nonviolent ways to bring about policy change, thus precluding the necessity condition from being met. But, as Pasternak correctly points out, this "underestimates the debilitating impact of pervasive socioeconomic and racial injustices."[13] The poverty experienced by many oppressed citizens makes it difficult to participate politically in multiple ways (sometimes even including the ability simply to cast votes in elections). Histories of oppression often involve the entrenchment and persistence of prejudicial views of oppressed citizens, silencing whatever political voices are able to make it into public discourses. And sometimes governments make policy decisions that overtly diminish the political power oppressed citizens are able to wield (e.g., through gerrymandering).

The upshot is that in some ostensibly democratic societies, one of the primary injustices political rioting responds to is the fact that the various nonviolent means of bringing about policy change are ripped out from under oppressed citizens' feet. Under such conditions, political rioting may conceivably be the only way for oppressed citizens to secure policy changes that eliminate or ameliorate the severe injustices they suffer.

3. PRIVILEGED CITIZENS AND THE SCOPE OF NECESSITY

Pasternak's discussion of the necessity condition limits itself to oppressed citizens. She explicitly notes that "in the case of political riots, it must be the case that the injustice the protesters face affects their own lives in ways that render

11 Pasternak, "Political Rioting," 400.

12 Pasternak, "Political Rioting," 401.

13 Pasternak, "Political Rioting," 401.

other forms of protest inaccessible to them."[14] The result is that privileged citizens lack moral license to participate in political riots, even ones that respond to gross injustices. On the surface, this makes sense: in the kinds of democratic societies Pasternak's account focuses on, presumably, such privileged citizens have access to less violent means to seek the policy changes that riots aim to achieve. By virtue of their being privileged citizens, they are typically much better off economically, possess a fully respected political voice, and lack trouble casting votes in elections. As such, it would seem that the circumstances of privileged citizens render it impossible for them to satisfy the necessity condition.

This is where I disagree with Pasternak and believe her account could be substantially improved. It can sometimes be the case that participation of privileged citizens in otherwise justifiable political rioting can indeed satisfy the necessity condition and, in doing so, bolster the rioting's chances of success. I, therefore, submit that we should modify the scope of Pasternak's necessity condition and be willing to acknowledge the permissibility of privileged citizens rioting.

Pasternak's principal objection to privileged citizens rioting is that they have alternative options available to them that are closed off to oppressed citizens. While it might be true that privileged citizens as a class have options available to them that oppressed citizens as a class do not, it may be the case that not enough individual privileged citizens are willing to avail themselves of the alternative options to successfully ameliorate the injustice via less violent means.[15] To be sure, one can easily imagine a case where there just are not enough privileged citizens seeking the kinds of change required to eliminate or ameliorate the injustice to which rioting may be a permissible response, and this fact is an important part of why the kinds of protests that can lead to riots begin to emerge in response to the injustice.

If privileged citizens using their political power do not have a reasonable chance of success, most likely because a critical mass of fellow privileged citizens fails to see the need to change policies, then such alternative means are not substantively available to the privileged citizens cognizant of the need for change, undermining Pasternak's claim to the contrary. As a result, when a privileged citizen deliberates about whether she, as a conscientious individual, ought to participate in some riot that may unfold, she should only be required to seriously consider those alternative options that she reasonably believes could be successful in ameliorating the injustice. If it is clear that, say, waiting months or years to cast a single vote in a blood-red state for the progressive candidate and that beseeching fellow privileged citizens to do the same simply

14 Pasternak, "Political Rioting," 403n78.

15 Indeed, this is often precisely why systemic injustice persists.

will not be enough to affect the kinds of change required to ameliorate injustice, then there is a morally significant sense in which such alternative options are not plausibly available. As such, it might genuinely be the case that the participation of privileged citizens is necessary despite the theoretical availability of alternative options that are closed off to oppressed citizens. The point here is that, at least in cases where the injustice is entrenched, it is very unlikely that the few privileged citizens willing to resist the injustice will succeed by less violent means. Joining the oppressed citizens in rioting, that is, may be the only course of action reasonably open to privileged citizens.

An additional reason to support expanding the necessity condition to permit privileged citizens to riot on behalf of and alongside oppressed communities concerns the importance of eliminating or ameliorating the injustices to which the rioting responds. The stakes for oppressed communities are extremely high. The kind of severe oppression that might warrant political rioting often results in the deaths of innocent members of society. Through things like police brutality based on racist social norms, entrenched impoverishment forcing many members of oppressed communities to resort to criminal behavior to survive, or even, more mundanely, the lack of federal aid programs resulting in significantly worse health outcomes for members of oppressed communities, many oppressed citizens' lives are lost or severely impacted by the injustices to which rioting responds. This makes the need to eliminate or at least begin ameliorating such injustices an urgent moral imperative. If it is the case that participation of privileged citizens in rioting can improve the chances of success, then their participation is therefore supremely morally important—important enough, I submit, that an account of permissible rioting ought to accommodate such possibilities.

This ties closely to my understanding of the necessity condition. It will often be the case, I think, that such participation will increase the chances of successfully eliminating or ameliorating the injustices to which rioting responds. Given the political alienation that oppressed communities often suffer—the "sense of powerlessness, or the lack of belief in one's capacity to bring about change via the standard channels"—there will almost certainly be a large proportion of the population that dismisses oppressed rioters as mere criminals because the privileged citizens that dismiss their concerns are incapable of understanding the alienation and injustices against which their rioting justifiably lashes out.[16] But when privileged citizens see fellow privileged citizens rioting on behalf of and in solidarity with oppressed citizens, it seems plausible that at least some

16 Pasternak, "Political Rioting," 402. I would like to thank Donald Wagner for insightful discussion on this point.

heretofore unconvinced privileged citizens would come to recognize the need for change. It is not difficult to imagine a situation in which media coverage depicting privileged citizens rioting alongside and in solidarity with oppressed citizens has a positive impact on the attitudes and beliefs of ambivalent privileged citizens regarding the urgent need for change, thus increasing the chances of success. This is admittedly speculative. But it seems much more likely than privileged citizens' rioting *reducing* the likelihood of success. The upshot is that if rioting is morally permissible for oppressed citizens, it may also be the case that rioting is equally morally permissible for privileged citizens.

4. Conclusion

I have argued, contra Pasternak's suggestion, that an account of permissible political rioting should include the possibility that privileged citizens may permissibly take part in at least some otherwise justified riots. I have done so on two grounds: (1) at least in cases of entrenched injustice, it seems likely that rioting will be genuinely necessary for privileged citizens every bit as much as it is necessary for oppressed citizens; and (2) given the political alienation experienced by oppressed communities that are justified in rioting, permitting the participation of privileged citizens in riots will likely increase the chances of success. I have ultimately advanced a very narrow argument: *only* insofar as rioting by oppressed citizens specifically in response to gross injustice is permissible, it *may* be permissible for privileged citizens to participate in the rioting alongside and in solidarity with oppressed citizens.

An important upshot of my argument is that there will almost always be an important affinity between political rioting that is justified on Pasternak's modified account and more conventional civil disobedience, which is typically understood not to admit of violence.[17] Many theorists understand civil disobedience to be inherently nonviolent because of the fact that it should express an inherent respect for the authority of the state. Such respect for authority involves a public commitment to realizing the ideals of the shared democratic project. I take it that this is exactly what is expressed in cases where oppressed and privileged citizens riot alongside each other: the rioting I have in mind occurs only because it has become necessary to remind the democratic society of the ideals and shared political project to which it is ostensibly collectively committed but is failing to realize. Both civil disobedients and justified rioters, therefore, express a similar commitment to and demand for just social

17 Rawls, *A Theory of Justice*, 336–37; Lefkowitz, "On a Moral Right to Civil Disobedience," 216; Brownlee, *Conscience and Conviction*, 29–46.

conditions.[18] Expanding the scope of our necessity condition, then, both improves Pasternak's account and helps show that justified instances of rioting are morally closer to traditional civil disobedience than many people seem willing to concede. This affinity should result in less condemnatory responses to political rioting.

Duke University
thomas.carnes@duke.edu

REFERENCES

Brownlee, Kimberley. *Conscience and Conviction: The Case for Civil Disobedience.* Oxford: Oxford University Press, 2012.

Lefkowitz, David. "On a Moral Right to Civil Disobedience." *Ethics* 117, no. 2 (January 2007): 202–33.

Pasternak, Avia. "Political Rioting: A Moral Assessment." *Philosophy and Public Affairs* 46, no. 4 (Fall 2018): 384–418.

Rawls, John. *A Theory of Justice.* Rev. ed. Cambridge, MA: Harvard University Press, 1999.

Waddington, David. "The Madness of the Mob? Explaining the 'Irrationality' and Destructiveness of Crowd Violence." *Sociology Compass* 2, no. 2 (March 2008): 675–87.

18 Pasternak, "Political Rioting," 395–96.

Journal of Ethics and Social Philosophy
VOL. 26, NO. 3 · FEBRUARY 2024
https://doi.org/10.26556/jesp.v26i3.3233

GASLIGHTING AND PEER DISAGREEMENT

Scott Hill

According to the Dilemmatic Theory proposed by Kirk-Giannini: a subject, S_1, gaslights another subject, S_2, with respect to a proposition, p, iff (1) S_1 intentionally communicates p to S_2; (2) S_2 knows (and S_1 is in a position to know) that if p is true, then S_2 has good reason to believe she lacks basic epistemic competence in some domain, D; (3) S_1 does not correctly and with knowledge-level doxastic justification believe p, and S_1 does not correctly and with knowledge-level doxastic justification believe that S_2 lacks basic epistemic competence in D; and (4) S_2 assigns significant weight to S_1's testimony.[1]

Part of what sets this theory apart is that it is not supposed to include any appeal to social hierarchies or testimonial injustice or the intentions of the gaslighter (other than the intention to communicate p). At the same time, it articulates and makes explicit a feature of gaslighting that, in retrospect, is clearly central but, until now, has gone largely unrecognized.[2] In particular, the theory illuminates the distinctive dilemmatic structure of gaslighting. This kind of insight, something that in retrospect seems like it should have been obvious and central all along, is the mark of an important contribution.

The theory also delivers the judgment that gaslighting occurs in the following cases:

> *Central Case*: Gregory seeks to rob Paula of her aunt's jewels, which are hidden in her attic. He routinely searches the attic, at which times the sound of his footsteps and the dimming of the house's gaslights are clearly perceptible to Paula. But when Paula discusses her observations with Gregory, he insists that she is merely imagining the footsteps and dimmings. Distressed, Paula begins to fear that she is losing her sanity.

1 Kirk-Giannini, "Dilemmatic Gaslighting," 757.

2 Kirk-Giannini gives credit where credit is due, however. He points out that Spear ("Gaslighting, Confabulation, and Epistemic Innocence") briefly touches on a similar idea. And he explicitly identifies elements of his theory influenced by Ivy ("Gaslighting as Epistemic Violence") and Podosky ("Gaslighting First- and Second-Order"). He also notes that he draws on and builds his theory in part out of examples first introduced by Abramson ("Turning Up the Lights on Gaslighting").

> *Skeptical Peers*: I moved out of one field of philosophy in grad school due to an overwhelming accumulation of small incidents. . . . When I tried to describe to fellow grad students why I felt ostracized or ignored because of my gender, they would ask for examples. I would provide examples, and they would proceed through each example to "demonstrate" why I had actually misinterpreted or overreacted to what was actually going on.[3]

Kirk-Giannini shows that the Dilemmatic Theory accommodates intuitions about a wide variety of cases, including variants of the above. And he shows that more traditional theories have trouble accommodating these cases.

Nevertheless, I think there are variants of Skeptical Peers that may be cause to modify the Dilemmatic Theory. Consider:

> *Skeptical Peers II*: Paula tells her peers that she feels ostracized and ignored in her subfield of philosophy because she is a woman. Paula provides examples to illustrate. When Paula considers the examples, they seem to her to clearly be cases that illustrate discrimination. When her peers consider the cases, they seem to them to clearly not be such cases. Paula forms her belief on the basis of her personal experiences. Paula's peers form their belief on the basis of statistical reasoning about her descriptions of the case. Paula and her peers assign significant weight to each other's testimony.

If we stipulate that Paula's peers do not correctly believe that she is mistaken, then the theory has the result that Paula's peers gaslight her. That is not the basis of an objection. The question of whether gaslighting can occur in the absence of intention is a matter of dispute in the literature.

I want to focus on a different seeming result of the theory. At first glance, it might seem that the Dilemmatic Theory has the additional result that gaslighting can go in either direction in this case. If Paula is right or if her peers are not justified in believing that she is wrong, then Paula's peers gaslight her. And, if Paula is wrong or if she is not justified in believing that her peers are wrong, then Paula gaslights her peers. In the latter case, condition 1 is satisfied because Paula testifies to her peers that she is ostracized and ignored because she is a woman. Condition 2 is satisfied because if Paula is right, then her peers lack basic epistemic competence in assessing examples of discrimination. Condition 4 is satisfied because Paula's peers assign significant weight to Paula's testimony.

3 Jender, "But the Women Never Say Anything Interesting," as cited in Abramson, "Turning Up the Lights on Gaslighting," 5.

Condition 3 *seems* to be satisfied. There are two ways in which the case can be formulated so that condition 3 might appear to be satisfied. One way 3 might be satisfied is simple. If Paula is wrong and she was not discriminated against, then the condition is satisfied because she does not correctly believe her peers lack the relevant basic epistemic competence.

The other way 3 might be satisfied is a bit more complicated. Suppose Paula is right, and she was discriminated against, but she does not believe it. Stipulate that the disagreement with her peers causes her to be so shaken and distressed that she becomes agnostic and does not believe her peers lack the relevant basic epistemic competence, and she does not believe that she has been discriminated against. Nevertheless, she thinks it is worthwhile to present her case. This could be because she feels defensive. Or it could be because she believes in intellectual diversity, and so although she does not believe what seems to her to be true, she thinks it is important to get her different perspective on the table in discussion with her friends. We can imagine something similar happening in Central Case. Paula might be so shaken by Gregory's testimony that she no longer believes the gaslights flickered. But she may still feel compelled to assert that the lights have flickered. This could be because she is feeling defensive or because she thinks, even though she may well be wrong, her testimony and perspective should be heard as one voice in the conversation.

So the theory, either because Paula is wrong or because she is right but has been shaken by disagreement, seems to have the implication that Paula gaslights her peers.

Either way, the two main camps in the literature would be uneasy with this result. One camp would be uneasy because they take Paula to lack the intentions required for gaslighting. The other camp would be uneasy because they take gaslighting to occur only in the direction of more to less powerful people. Paula is less powerful than her peers. So she does not gaslight. So this result, if Kirk-Giannini were to accept it, would put him outside of the mainstream.

Being outside the mainstream may not be bad in itself. But if one's theory seems to depart from the mainstream, then it is important to either give a story about why it turns out to be acceptable to depart from the mainstream or give a story about why the theory does not really deliver the relevant out of the mainstream judgment.

In the present case, Kirk-Giannini may plausibly reject the claim that his theory has the relevant result. In particular, he may note that there is an asymmetry between Paula and her peers. In Skeptical Peers II, Paula is not calling into question a basic epistemic competence. She is instead calling into question an advanced epistemic competence. She calls into question the ability of her peers to evaluate complicated statistical claims. Paula's peers form their belief

based on advanced statistical reasoning. Paula forms her belief based on her experience that comes from her position of marginalization. Advanced statistical reasoning is not a basic epistemic competence. As Kirk-Giannini puts it:

> There are some domains in which our beliefs are not plausibly regarded as formed on the basis of any basic epistemic competence. First, there are beliefs about theoretical domains like advanced mathematics, the natural and social sciences, philosophy.... Second, there are beliefs which ... are formed on the basis of evidence which is subtle or otherwise difficult to interpret.[4]

Indeed, given that the report in the original Skeptical Peers is that the grad student peers "proceed through" the examples and "demonstrate" that she is mistaken, it sounds like they are employing an advanced rather than basic epistemic competence. On the other hand, experience that comes from one's position of marginalization, one might maintain, is a basic epistemic competence. So condition 3 is unsatisfied. Paula's peers gaslight her. But Paula does not, given the Dilemmatic Theory, gaslight her peers. And Kirk-Giannini has a plausible way of resisting the argument I gave above.

So far so good. But if one takes this line, then it seems to me the theory is subject to a different counterexample. Consider:

> *Skeptical Peers III*: Paula tells her peers that she feels ostracized and ignored in her subfield of philosophy because she is a woman. Paula provides examples to illustrate. She evaluates those examples via her views about complicated statistical inferences, sociological background claims, and philosophical reflection about how women in philosophy are generally treated. Her peers know that she is right. But they dismiss her concerns as being based on a misunderstanding of complicated statistics. They tell her that because she is a woman, she is incapable of competently engaging in the kind of advanced statistical reasoning required to understand the examples. They maintain that while women have all basic epistemic competences, they do not have the advanced epistemic competences that are unique to men. Distressed, Paula begins to wonder whether they might be right. And she thinks she might be misunderstanding the complicated statistics and, therefore, whether she has been discriminated against.

If the Dilemmatic Theory is combined with the view that advanced statistical reasoning is not a basic competence, then the theory delivers the result that

4 Kirk-Giannini, "Dilemmatic Gaslighting," 765.

Paula's peers do not gaslight her. In order to satisfy condition 3, Paula's peers must call into question a basic epistemic competence. But in this case, they do not. They instead cast doubt on whether she is competent in advanced statistics because, they claim, women are incapable of doing advanced statistics. And yet, this seems like a paradigm example of gaslighting.

Let me say more to defend my judgment that Paula's peers gaslight her in Skeptical Peers III. Note that this variant is merely a way of filling in the details of Skeptical Peers. As Kirk-Giannini notes, Skeptical Peers first appeared on the blog *What Is It Like to Be a Woman in Philosophy?* and then was adopted by Abramson in her list of eight central cases of gaslighting out of which she builds her theory. Kirk-Giannini observes that the case is underspecified in various ways. And yet, even without very many details being filled in, it is nevertheless a paradigm example of gaslighting. Our reaction is that it is a case of gaslighting. Our reaction is not that we need to hear more from the woman reporting her experience before we can tell whether it is really gaslighting. And Kirk-Giannini points out that one of the details missing from the case is whether the woman's peers are acting with the intention Abramson thinks is required for gaslighting (the intention to subvert or control). Kirk-Giannini reasons that this suggests that whether intention occurs in the case is irrelevant to whether gaslighting occurs. Kirk-Giannini puts it this way:

> The case as Abramson presents it is underspecified: it does not tell us anything about the intentions of the fellow graduate students. . . . We can imagine that the perpetrators of the gaslighting in Skeptical Peers do indeed have the kinds of subterranean motivations Abramson regards as individuative of gaslighting. But we can also imagine that they do not. . . . The fact that we can identify Skeptical Peers as a case of gaslighting without knowing about the intentions of the gaslighters suggests that our judgment about the case is not sensitive to facts about those intentions. This conclusion is further suggested by the observation that our intuitive sense that the victim's fellow graduate students are gaslighting her persists when we fill out the case so that they lack an intention to subvert or control her. If this line of argument is sound, it must be possible for there to be gaslighting in the absence of the psychological features Abramson and other intentionalists identify, common or salient though those features may be.[5]

I think we can say the same thing about the lack of details in Skeptical Peers concerning exactly what kind of competence is being called into question.

5 Kirk-Giannini, "Dilemmatic Gaslighting," 750–51.

There are no details in the original Skeptical Peers about whether what is called into question is the graduate student's knowledge from a position of marginalization or her ability to do complicated statistics or anything else. If we follow Kirk-Giannini's reasoning, this suggests that exactly which epistemic competence is called into question is not relevant to our intuitions about whether she is gaslighted. Think about it this way: suppose the woman who wrote the blog post on *What Is It Like to Be a Woman in Philosophy?* comes back to fill in the details and reveals that she was dismissed by her peers for her alleged lack of competence in advanced statistics on the basis of being a woman.[6] We would not then conclude that she is mistaken and that her peers did not gaslight her.

Furthermore, Skeptical Peers is an especially central example for testing theories of gaslighting. As Kirk-Giannini puts it:

> There is thus an important dialectical difference between cases like Bird and Bill and cases like Skeptical Peers. Whereas existing accounts' difficulties with capturing the intuition that certain versions of Skeptical Peers involve gaslighting give us reason to hope for an account which does better, the fact that (Dilemmatic Gaslighting) classifies certain versions of Bird and Bill as gaslighting does not indicate that it struggles to capture our intuitions in the same way.[7]

So Kirk-Giannini takes it to be especially important to match intuition in Skeptical Peers. And there are ways of filling in the details of Skeptical Peers in which our intuitions do not change but in which the Dilemmatic Theory seems to give a counterintuitive result. If we follow Kirk-Giannini's reasoning here, then it seems that the point he makes about others' theories also applies to his theory. It is a serious problem if the theory diverges from intuitions about Skeptical Peers III.[8]

University of Innsbruck
hillscottandrew@gmail.com

6 Jender, "But the Women Never Say Anything Interesting."

7 Kirk-Giannini, "Dilemmatic Gaslighting," 768.

8 For comments and discussion, I thank Richard Greene, Dale Miller, Michelle Lynn Panchuk, Lewis Michael Powell, Jack Slate, and Sameer Yadav. Special thanks to the referees for *JESP*. Extra special thanks to Cameron Kirk-Giannini for detailed comments on several drafts. Work on this paper was funded as part of the Euregio Interregional Project Network IPN 175 "Resilient Beliefs: Religion and Beyond."

REFERENCES

Abramson, Kate. "Turning Up the Lights on Gaslighting." *Philosophical Perspectives* 28, no. 1 (December 2014): 1–30.

Ivy, Veronica. (McKinnon, Rachel.) "Gaslighting as Epistemic Violence: 'Allies,' Mobbing, and Complex Posttraumatic Stress Disorder, Including a Case Study of Harassment of Transgender Women in Sport." In *Overcoming Epistemic Injustice: Social and Psychological Perspectives*, edited by Benjamin R. Sherman and Stacey Goguen, 285–322. London: Rowman and Littlefield, 2019.

Jender. "But the Women Never Say Anything Interesting." *What Is It Like to Be a Woman in Philosophy?* October 29, 2010. https://beingawomaninphilosophy.wordpress.com/2010/10/29/but-the-women-never-say-anything-interesting/.

Kirk-Giannini, Cameron Domenico. "Dilemmatic Gaslighting." *Philosophical Studies* 180, no. 3 (March 2023): 745–72.

Podosky, Paul-Mikhail Catapang. "Gaslighting, First- and Second-Order." *Hypatia* 36, no. 1 (December 2021) 207–27.

Sherman, Benjamin R., and Stacey Goguen, eds. *Overcoming Epistemic Injustice: Social and Psychological Perspectives*. London: Rowman and Littlefield, 2019.

Spear, Andrew D. "Gaslighting, Confabulation, and Epistemic Innocence" *Topoi* 39, no. 1 (February 2020): 229–41.

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