JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY

VOLUME XXX · NUMBER 2
*April 2025*

ARTICLES



DISCUSSION



OPEN ACCESS FREE FOR AUTHORS AND READERS ISSN 1559-3061

JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY

VOLUME XXX · NUMBER 2
*April 2025*

ARTICLES



DISCUSSION

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.

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

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

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

*Editors-in-Chief*
Sarah Paul
Matthew Silverstein

*Associate Editors*
Rima Basu
Saba Bazargan-Forward
Brian Berkey
Ben Bramble
James Dreier
Julia Driver
Alex Gregory
Sean Ingham
Anthony Laden
Coleen Macnamara
Elinor Mason
Simon Căbulea May
Tristram McPherson
Hille Paakkunainen
David Plunkett
Sam Shpall
Kevin Toh
Mark van Roojen
Han van Wietmarschen
Jonathan Way
Vanessa Wills

*Symposia Editor*
Errol Lord

*Managing Editor*
Chico Park

*Copyeditor*
Lisa Y. Gourd

*Proofreader*
Susan Wampler

*Typesetter*
Matthew Silverstein

*Editorial Board*
Elizabeth Anderson
David Brink
John Broome
Joshua Cohen
Jonathan Dancy
John Finnis
Leslie Green
Karen Jones
Frances Kamm
Will Kymlicka
Matthew Liao
Kasper Lippert-Rasmussen
Stephen Perry
Philip Pettit
Gerald Postema
Henry Richardson
Thomas M. Scanlon
Tamar Schapiro
David Schmidtz
Russ Shafer-Landau
Tommie Shelby
Sarah Stroud
Valerie Tiberius
Peter Vallentyne
Gary Watson
Kit Wellman
Susan Wolf

*Journal of Ethics and Social Philosophy*
VOL. 30, NO. 2 · APRIL 2025
*https://doi.org/10.26556/jesp.v30i2.4414*
© 2025 Author

# DEMOCRACY, DISOBEDIENCE, AND ACCOUNTABILITY

*Henry Krahn*

PROTESTORS tend to impose on others. When Extinction Rebellion activists glue themselves to the road to challenge climate policy, it makes people late for work. When students pitch tents on university campuses to protest the ongoing conflict in Gaza, they prevent people from using public spaces. When protestors burned down a Minneapolis police station over the murder of George Floyd, they destroyed a piece of public property. In these ways, protest of all kinds—violent and nonviolent, organized and disorganized, civil and uncivil—can interfere with the interests of others. Protest, let us say, is often *burdensome*.[1]

This burdensome aspect of protest is typically intentional. Protestors do not glue themselves to a road by accident. Nor is the intentional imposition of burdens a new feature of protest. To the contrary, it is deeply ingrained in our political culture. Many protestors believe that some degree of social disruption is essential to their cause. Martin Luther King Jr., for example, wrote that ethical appeals in protest must be "undergirded by some form of constructive coercive power."[2] Still, the fact that protest can be inconvenient, disruptive, and destructive is a perennial cause of political opposition to protest. Even sympathizers may find their sympathies tested when they witness the destruction of public property or find themselves stuck in traffic behind a blockade. Minimally, interference of this kind in the interests of others seems to call for some form of justification.

A common line of thought among philosophers writing on civil disobedience has been that this is because burdensome protest is *undemocratic*. When protestors impose burdens on others, the thought goes, they attempt to directly bring about their desired political outcome rather than leaving the issue to the majority. As John Rawls, the leading liberal theorist of civil disobedience, puts it, civil disobedience "tries to avoid the use of violence, especially against

1 I use 'protest' here as a catchall term that includes, for example, marches, demonstrations, sit-ins, and phone zaps, as well as whistleblowing, political graffiti, hacktivism, and targeted property destruction.

2 King, *Where Do We Go from Here?* 137.

persons, not from the abhorrence of the use of force in principle, but because [civil disobedience] is a final expression of one's case. . . . Civil disobedience is giving voice to conscientious and deeply held convictions; while it may warn and admonish, it is not itself a threat."[3] Because this objection is levelled against protest that is described as violent or forceful, call it the *democratic objection to force*, or just the *democratic objection*.[4]

As stated, the democratic objection is clearly overbroad. Practically all civil disobedience involves tactics that might be described as forceful or violent, to say nothing of other kinds of protest.[5] In some jurisdictions, simply sitting on a road may constitute "violent coercion."[6] For this reason, most authors who endorse the democratic objection carve out a special justification for certain incidental uses of force or violence. Others draw a more radical conclusion, seeking to reject the democratic objection altogether. But there is an important question in this dialectic that has received regrettably little attention: Is it true that force and violence per se imply an attempt to directly bring about a political outcome? If not, then we can take the democratic objection seriously without subscribing to a narrow view of the justification of forceful and violent protest.

This essay offers a response to the democratic objection that adopts this strategy. I argue that we can distinguish some burdensome protest from attempts to directly bring about a desired political outcome, avoiding the democratic objection to force. To this end, I propose that protest is sometimes a form of holding others accountable. When we hold others accountable, we are often engaged neither in persuasion nor in an attempt to force our views on others. Rather, we are engaged in a form of communication that uses the imposition of burdens to make a moral appeal to others. If burdensome protest can be interpreted in similar terms, then protestors may *use force* without *forcing* their beliefs on others.

I begin in section 1 by unpacking the democratic objection to force in more detail. In sections 2 and 3, I set out an account of holding others accountable and argue that protest, interpreted in these terms, can avoid the democratic

3 Rawls, "Definition and Justification of Civil Disobedience," 106.

4 Even some proponents of forceful and violent protest accept this point. See Aitchison, "Coercion, Resistance and the Radical Side of Non-Violent Action"; and Livingston, "Non-violence and the Coercive Turn." For a general discussion of worries about coercion, see also Delmas and Brownlee, "Civil Disobedience."

5 Morreall, "The Justifiability of Violent Civil Disobedience," 136. Morreall's precise word here is 'coercion', but nothing is lost in the rephrasing. See also Greenawalt, "Justifying Nonviolent Disobedience," 179; and Moraro, "Violent Civil Disobedience and Willingness to Accept Punishment," 274.

6 I have in mind here the German Penal Code. For discussion, see Celikates, "Rethinking Civil Disobedience as a Practice of Contestation," 41–42.

objection to force. In section 4, I defend this interpretation of protest. I conclude in section 5 by considering the scope of the paper's conclusions.

## 1. THE DEMOCRATIC OBJECTION TO FORCE

Let me start by laying out the democratic objection in general terms and motivating my approach to the issue.

The democratic objection to force is founded on the idea that democratic citizens have a duty to respect the outcomes of democratic decision-making. According to this idea, democratic citizens are entitled to participate in collective decision-making processes by doing things like voting, attending town hall meetings, writing editorials, and so on. But part of being a good democratic citizen is respecting democratic decisions even when one disagrees—being, in Locke's turn of phrase, "concluded by the majority."[7] This duty raises a problem for civil disobedience insofar as the law disobeyed is a product of democratic decision-making. Nevertheless, the argument usually goes, the duty does not require that citizens always and unfailingly adhere to the letter of the law. For even though civil disobedience involves disobeying the law, it does so in a way that demonstrates respect for the law. Unlike other kinds of lawbreaking, civil disobedience is nonviolent, conscientious, political, persuasive, and done with a willingness to accept punishment for one's actions. In contrast, the objection holds, forceful and violent disobedience involves violating the duty of respect for the law. Civil disobedience is a principled attempt to persuade the public; the use of violence or force is an attempt to directly bring about a desired political outcome regardless of what other citizens happen to think on the issue. Thus, forceful or violent disobedience is thought to be an attempt by a minority to force their views on the democratic majority, which is incompatible with respect for the outcomes of democratic decision-making.

This argument is worth taking seriously. The claim that democratic citizens have a duty to respect the law is a plausible one, with deep roots in political philosophy and many influential contemporary proponents.[8] Indeed, the literature on civil disobedience historically begins from an even stronger claim—that citizens have a *pro tanto* duty to obey the law.[9] And it is an accepted feature of the

7 See Locke, *Second Treatise of Government*, sec. 96.

8 Jeremy Waldron writes, "When something is enacted as law or as a source of the law, I believe it makes on us a demand not to immediately disparage it, or think of ways of nullifying it or getting around it. . . . [This] is a demand for a certain sort of recognition and, as I said, respect" (*Law and Disagreement*, 100). See also, e.g., Christiano, *The Constitution of Equality*, 250; and Stilz, *Liberal Loyalty*, 98.

9 Lai, "Justifying Uncivil Disobedience," 90.

common understanding of civil disobedience that it can demonstrate respect for the law.[10] The question, then, is whether force and violence are incompatible with respect for the outcomes of democratic decision-making. Certainly, many have thought so. As we have already seen, Rawls argues that civil disobedience cannot be violent because violence uses threats rather than political appeals to get its way.[11] Other significant authors in the civil disobedience literature like Ronald Dworkin, Peter Singer, David Lefkowitz, and Daniel Markovits make similar claims, even while they disagree with the substance of Rawls's view.[12]

To be sure, the democratic objection has taken many forms. Perhaps unsurprisingly, few who endorse the objection see it as a categorical rejection of force and violence. Instead, it is common to qualify the democratic objection so as to allow some *incidental* force and/or violence.[13] Markovits's view is representative. His account notes a key distinction between the *fact* and the *outcome* of democratic reengagement: while disobedients may justifiably use coercion to secure the fact of democratic reengagement, they may not coerce a specific political outcome.[14] William Smith likewise defends coercion to kick-start debate of a neglected issue in the public sphere, though not to secure the success of a particular position.[15] And Piero Moraro defends protest that uses force to get the full attention of one's democratic peers without forcing them to adopt a view.[16] Beyond these authors, this line of argument is well established in the literature.[17]

10 King, for example, writes in "Letter from a Birmingham City Jail" that, far from disrespecting the law, civil disobedience "in reality express[es] the very highest respect for law" (1121).

11 Rawls, "Definition and Justification of Civil Disobedience," 106.

12 Dworkin argues that nonpersuasive civil disobedience is harder to justify than persuasive civil disobedience because it attempts to make the majority's chosen policy more costly rather than change the majority's mind ("Civil Disobedience and Nuclear Protest"). Lefkowitz, citing Singer, argues that civil disobedience cannot be coercive because coercive disobedience threatens to usurp the "equal authority of all citizens to determine what the law ought to be" ("On a Moral Right to Civil Disobedience," 216). For his part, Singer argues that civil disobedience for "publicity purposes" must avoid violence because "to use violence is to obliterate the distinction between disobedience for the sake of publicity and disobedience designed to coerce or intimidate the majority" (*Democracy and Disobedience*, 82). Markovits, in contrast, allows for disobedients to use coercion to trigger democratic reengagement with an issue but not to force a specific political outcome ("Democratic Disobedience," 1941).

13 I thank an anonymous reviewer for pressing this point.

14 Markovits, "Democratic Disobedience," 1941.

15 Smith, "Civil Disobedience and the Public Sphere," 160–61.

16 Moraro, "Respecting Autonomy Through the Use of Force," 68.

17 King himself conceived of civil disobedience as using force to secure the conditions for negotiation. Kimberley Brownlee and Candice Delmas each cite and substantively agree

This incidental argument, however, yields only a narrow defense of force and violence. First, the argument applies only to cases of civil disobedience and, moreover, to cases of civil disobedience in which force or violence is used to secure the fact and not the outcome of democratic reengagement. Second, it is difficult to see how the use of force or violence to prompt democratic reconsideration of an issue can be disentangled from the use of force or violence to push for a specific outcome, especially when both are likely to be performed by activist groups with a clear political program. Third, this sort of argument presents force and violence as instrumental preliminaries to persuasion—noncommunicative measures designed to secure the conditions for civil disobedients to make their case. But as many have argued, force and violence can play an essential role in shaping the content of a protest's message.[18] These issues suggest that there is a range of (not incidentally) forceful and violent protest calling for independent theoretical attention.

Accordingly, some authors have sought to make more room for force and violence in our philosophical understanding of protest. Barbara LaBossiere, Jennifer Welchman, and A. John Simmons have argued that the rejection of violence by theorists of civil disobedience is ahistorical, and Andreas Marcou has made a legal case for the compatibility of violence with civil disobedience.[19] Others have drawn attention to the idealizing assumptions undergirding the democratic objection. The democratic objection is an objection to protest that

---

with Markovits regarding the democratic objection, though Delmas also takes the argument a step further to apply it to uncivil disobedience (Brownlee, *Conscience and Conviction*, 176; and Delmas, *A Duty to Resist*, 56–57). Lefkowitz takes a thicker, more moralized view of coercion. Accordingly, he contends that civil disobedience cannot be coercive, but it may be violent insofar as violence can be noncoercive—as in, for instance, the symbolic destruction of a monument ("On a Moral Right to Civil Disobedience," 216). Robin Celikates agrees that the use of nonpersuasive means can be put to democratic ends, though he does not appear to place such strong limits on civil disobedience ("Democratizing Civil Disobedience").

18 Robin Celikates claims that the communicative dimension of civil disobedience in fact *depends* on moments of "real confrontation," often including force or violence ("Democratizing Civil Disobedience," 988). Steve Coyne denies Rawls's claim that violence is incompatible with civil disobedience as a mode of address, writing, "Violence can often be a spectacularly powerful way of addressing someone" ("The Role of Civility in Political Disobedience," 224–25). Edmund Tweedy Flanigan defends the use of force on the grounds that it may be a fitting way of rejecting an attacker's refusal of moral regard ("Futile Resistance as Protest," 652–53). See also Kling and Mitchell, "Responsibility and Accountability"; and Marcou, "Violence, Communication, and Civil Disobedience."

19 LaBossiere, "When the Law Is Not One's Own"; Welchman, "Is Ecosabotage Civil Disobedience?"; Simmons, "Disobedience, Nonideal Theory, and Historical Illegitimacy"; and Marcou, "Violence, Communication, and Civil Disobedience." See also Smart, "Defining Civil Disobedience."

fails to show respect for democratic decision-making, so if a society's decision-making is undemocratic, then the objection loses much of its force.[20] Thus, Candice Delmas, for example, has argued that the democratic objection has limited application under present conditions of political marginalization.[21] However compelling, these arguments largely sidestep the normative issue at the heart of the democratic objection: whether force and violence are undemocratic *in principle*.[22]

The account I offer throughout the rest of this paper is designed to tackle this issue. I contend that the use of force and violence in protest need not be an attempt by protestors to force their views on others but may instead be a communicative effort compatible with respect for democratic decision-making.[23] Although the democratic objection has typically been considered by authors writing on civil disobedience, I engage with it as a broader problem for forceful and violent protest and offer a defense of protest that includes but is not limited to civil disobedience. In so doing, my goal is not to refute the democratic objection but to show that even if we take the democratic objection seriously, it may leave a good deal of force and violence on the table.

Before moving on, however, a note of clarification. I use the terms 'force' and 'violence' loosely in this essay to refer to a range of violent and nonviolent protest tactics. I will not attempt to define these terms, since they are used

20 A closely related point is that under conditions of sufficient injustice, there is no general duty to obey the law, and thus civil disobedience requires no special justification. See Lyons, "Moral Judgment, Historical Reality, and Civil Disobedience," 46; and Lefkowitz, "On a Moral Right to Civil Disobedience," 205, 209.

21 "To dismiss incivility as a threat to an otherwise stable democracy is most likely to assert stability (the kind that stems from a shared commitment to mutual reciprocity) where it has already been lost" (Delmas, *A Duty to Resist*, 56). See also Lai, "Justifying Uncivil Disobedience."

22 Simmons is an exception: "Even if we instead accept Rawls's ('public and political') requirements, it is simply not at all clear why violent acts could not be addressed to the public in the right way—as an attempt, say, to get the majority to reconsider its position on the justice of some policy" ("Disobedience, Nonideal Theory, and Historical Illegitimacy," 34–35). Although compelling, Simmons's remarks on this point are not developed at length.

23 This argumentative strategy is closest to Moraro's in "Respecting Autonomy Through the Use of Force." Unlike the argument offered in this paper, Moraro appears to offer a defense of the use of force in *civil disobedience* insofar as it serves the ends of a persuasive political appeal. But Moraro's argument resembles my own in that he argues that force need not involve an attempt to force others to do as one wishes and may instead be a part of one's political appeal. My argument can therefore be read as expanding on Moraro's conclusions by moving beyond civil disobedience and providing a thicker explanation of how force figures into the political message of protest.

loosely and in conflicting ways by many of the authors I discuss.[24] For our purposes, we need only note that these tactics impose burdens on their targets and are therefore thought to be undemocratic.

## 2. REPROOF

My defense of forceful and violent protest will depend on the notion of a form of holding others accountable called *reproof*. To introduce the idea, let us start by considering two more familiar responses to bad conduct: persuasion and sanctioning.

### *2.1. Persuasion and Sanctioning*

First, when others behave badly, we might attempt to persuade them to behave better.[25] We might engage them in a dialogue, offering reasons why their past conduct was bad and imploring them to reform. This is an option we are likely to take when delicacy is required, when the misconduct is relatively slight, or when we deal with friends and family. Suppose I lend my friend a book, and she carelessly drops it in a puddle. I might try to persuade her to be more careful in the future by arguing that she has reason to take better care of her friends' belongings. Or suppose I overhear my brother gossiping about me at a party. I might then go over and try to persuade him that he ought to respect my privacy.

As a response to misconduct, persuasion is motivated by a concern for the reasons of its target. In attempting to persuade my friend to take better care of others' belongings, I aim to have no influence over her conduct other than what might follow from the success of my arguments. I appeal to considerations she accepts as important in order to focus her attention on important facts that she has disregarded. Reasons come first, and the desired change in conduct is downstream. Thus, persuasion aims at a change in conduct *through* a change in the reasons endorsed by its subject. And it matters what reasons the subject ends up endorsing. In trying to persuade my friend to act differently, I aim for them to take up and act on precisely the reasons I present to them. So it will do me no good to get worked up and issue wild-eyed accusations. My persuasion will succeed only if my friend accepts the reasons I give her and not if she superficially changes her behavior to avoid confrontation.

24 Rawls uses 'force' and 'violence' interchangeably. Dworkin talks about *nonpersuasive disobedience*, by which he appears to mean something like *coercion*, as Markovits, Aitchison, and Livingston understand it.

25 I say 'attempt' because it is debatable whether persuasion or sanctioning count as holding others accountable in the fullest sense. But this should not affect the substance of the argument.

But the fact that persuasion is nonconfrontational can make it an odd response to some kinds of bad conduct. Persuasion may be appropriate when others are receptive and well intentioned. Often, however, others seem unlikely to respond well to our persuasive efforts. If someone cuts in front of me in line, it might be the case that they are unaware of the norms of queueing. But it is more probable that they are aware of those norms and simply do not care. My gossiping brother likely knows it is wrong to badmouth me behind my back, but he does it anyway—the fact that he gossips within earshot suggests that he has no compunction about it. Persuasion is ill suited to these contexts. Sometimes, it seems, bad conduct calls for something beyond a persuasive response.

Hence a second kind of response to bad conduct: sanctioning. When we hold others accountable, some moral philosophers argue, we impose some burdensome treatment, or sanction, on them that is designed to get them to change their conduct.[26] There is considerable disagreement about how sanctions are intended to effect this moral change, but for our purposes, I will take a simple view of sanctions according to which sanctions work through the creation of disincentives.[27] So when I sanction you for $\phi$-ing, I create a reason for you not to $\phi$, and the reason is that if you $\phi$, you will be subjected to something unpleasant. This unpleasantness might have to do with the burden of social censure, with the effect on your welfare of being scolded, or with some other impact on you.

To fix ideas, consider some examples. If you like to tell offensive jokes in public settings, I might try to get you to stop by haranguing you every time you tell one. And if I have given up on convincing you, I might rely on the sheer unpleasantness of my haranguing. Likewise, if I am having a conversation with someone who frequently interrupts me, I might decide that every time they interrupt me, I will interrupt them back. If I am a parent and hear that my children are starting to pick up swear words, I might make them deposit a quarter in a swear jar every time they swear. All these cases count as sanctioning for our purposes.

26 Sanctioning is an important element of many accounts in the Strawsonian tradition of moral philosophy. See Strawson, "Freedom and Resentment"; Macnamara, "Holding Others Responsible" and "Taking Demands Out of Blame"; Shoemaker, "Attributability, Answerability, and Accountability"; Wallace, *Responsibility and the Moral Sentiments*; and Watson, "Two Faces of Responsibility."

27 This is a commonplace though not uncontroversial understanding of sanctions. Many of the authors mentioned above would dispute it, but my aim here is to draw a broad distinction between everyday attempts to seek accountability, not to offer an analysis of the term 'sanction'.

As these examples suggest, sanctioning differs from persuasion in more than just its means. To persuade someone, I must have a certain concern for their reasons, but when sanctioning them, I may be indifferent to their reasons. When I interrupt the interrupter, for example, it is possible that they will come to realize that they have behaved rudely; but it is just as possible that they will stop interrupting me because they too dislike interruptions. To sanction the interrupter, I need not aim at one or another of these outcomes; I need aim only for the interrupter to stop interrupting me. In general, sanctions create motives for a change in conduct that are distinct from the reasons the conduct is bad. The fact that being interrupted is unpleasant is different from the fact that interrupting me is rude. As a result, the considerations in virtue of which persuasion is constrained and nonconfrontational do not apply to sanctioning. There is no need for sanctioning to be nonconfrontational because sanctioning does not depend on an appeal to its target's reason. While sanctioning might depend on its target being a rational agent—one who is responsive to the prudential impact of a sanction—it does not engage with its target as a reasoning agent.

The fact that sanctioning makes no appeal to its target's reason helps to explain why, like persuasion, sanctioning is also sometimes an odd response to bad conduct. Since the unpleasantness of a sanction is distinct from the reason why the conduct was bad, sanctions may be ill suited to bringing about swift and lasting change. If I use sanctions to get you to stop telling offensive jokes, you may not be able to realize what it is about your jokes that is offensive to me. And even if I am successful, I have given you a reason not to tell those jokes *only when I am around*.

### *2.2. The Concept of Reproof*

Sometimes when others act badly, we attempt neither to persuade nor to sanction them. This third option I call *reproof*, using the term in a special sense to pick out a response to bad conduct that falls somewhere between persuasion and sanctioning.[28] Reproof resembles sanctioning in that it works by impos-

28 My choice of this term and the subsequent discussion is strongly influenced by Coleen Macnamara's 2011 account of reproof in "On Holding Others Responsible," which itself builds on Antony Duff's 1986 treatment of the subject in *Trials and Punishments*. Duff's view is that reproof (Macnamara's term) is a kind of moral argument that aims to persuade a wrongdoer that they have acted badly so that they will be pained by their past conduct and thus commit to reform. The imposition of burdens can assist in this process, but they are not necessary for it. Macnamara, by contrast, distinguishes reproof from moral persuasion and insists that the imposition of burdens is an essential element of reproof. I agree with Macnamara that reproof is different from moral persuasion, in part because it involves the imposition of burdens. But I take Macnamara's 2021 account of reproof to lack some of the communicative features that Duff is at pains to emphasize. That is, if reproof achieves a change in conduct

ing burdens on others. Unlike sanctioning, however, reproof does not simply incentivize its targets to alter their behavior. In reproof, the imposition of burdens is intended and designed to communicate a reason why the target ought to behave differently. Reproof is thus both communicative, like persuasion, and burdensome for its target, like sanctioning.

To illustrate, imagine that you and I are part of a community cleanup of a local park. You finish the chocolate bar you are eating and drop the wrapper on the ground. In response, I turn to you, shocked, and say, "Don't do that!" In this example, I am not trying to *persuade* you not to litter. I have given you no argument to act otherwise. Rather, I am imposing a burden on you for your actions—the burden of dealing with my shocked disapproval. But neither am I using sanctions to incentivize you to change your conduct. Instead, it seems more plausible to say that I am trying, through the imposition of a burden, to get you to take seriously the fact that you should not litter. I am communicating with you, even if I am not trying to persuade you. To take a more extreme example, suppose Sam and Pat are in a relationship, and Pat finds out that Sam is cheating on them. In a fit of jealous rage, Pat carves the word 'cheater' into the paint of Sam's car. Here, Pat is not trying to persuade Sam, but Pat's actions have a distinct communicative purpose that goes beyond the creation of an incentive. Pat's actions seem to communicate accusation and condemnation in a way that persuasion cannot.

As the last example illustrates, reproof may resemble punishment in at least one important respect. Joel Feinberg argues that punishments are "conventional symbols of public reprobation."[29] To say this, Feinberg claims, is to say something such as champagne is symbolic of celebration or wearing black is symbolic of mourning. We might add knighting ceremonies or heckling a comedian to this list, in which the tap of a sword or the insult of a performer can take on an additional significance in virtue of social convention. Christopher Bennett and Michael McKenna contend that something similar is true of holding others accountable more generally.[30] As Bennett puts it, "the behavior which is the characteristic expression of blame has a purposive as well as an expressive or symbolic side. On the one hand, it symbolizes the wrongdoer's alienation [from other members of the moral community]; on the other, it serves to communi-

simply through the imposition of burdens, it is difficult to see how those burdens could lead a wrongdoer to accept that they have done wrong and commit to reforming. My discussion of the symbolic dimension of reproof throughout this section is intended to help fill this gap. See Duff, *Trials and Punishments*, 47–60; and Macnamara, "Holding Others Responsible," 90. See also Duff, *Punishment, Communication, and Community*.

29 Feinberg, "The Expressive Function of Punishment," 402.

30 See Bennett, "The Varieties of Retributive Experience"; and McKenna, *Conversation and Responsibility*.

cate that alienation to the wrongdoer."[31] This general point about accountability applies readily to the case of reproof. When Pat carves 'cheater' into the side of Sam's car in the above example, Pat imposes a burden on Sam with a meaning that goes beyond simply calling Sam a cheater. The damage done to Sam's car stands for Sam's alienation and communicates it to Sam.

Crucially, however, reproof does not just symbolize alienation; it also makes a moral appeal, providing its target with a reason to change their conduct. In the earlier example, when I tell you not to litter, I am not indifferent to your reasons. Given the context of your littering—a community cleanup—my reproof is intended to remind you of what we are doing and of reasons against littering that you presumably endorse. The burdens I impose on you do not just stand for alienation but also for oughts that you have disregarded. And it is an important part of what I do that I aim for you to recognize these oughts. Something would clearly go wrong if, after my reproof, you waited until I turned around and then threw your wrapper on the ground anyway. For I am reproving you not just because *I* see it as wrong to litter and want you to stop; I am telling you not to litter because I want *you* to see it as wrong and to stop accordingly. My reproof will fail if you treat it as a temporary inconvenience or an incentive to be mitigated. Like persuasion, reproof works through a change in its target's reasons, and it fails when no change occurs. Of course, I do not mean to suggest that attempts at reproof will always succeed purely by changing the mind of its target. The targets of such attempts may end up changing their conduct partly to avoid the unpleasantness of future burdens and partly because they see merit in the reproof. But affecting the target's cost-benefit calculations is not the point of reproof, and reproof will be successful in the fullest sense only if it achieves uptake of the reason the target's past conduct was wrong.

This last point about reproof allows us to see how the concept crystallizes some important ideas about accountability. As mentioned earlier, many philosophers think that holding others accountable involves burdensome treatment, which manifests in claims about sanctioning. But this claim sits uneasily with another common claim: that when we hold others accountable, we engage them in *moral address*, involving a genuine exchange of moral reasons.[32] The

31 Bennett, "The Varieties of Retributive Experience," 152.

32 Shoemaker, "Attributability, Answerability, and Accountability," 71. The language of moral address itself comes from Gary Watson. Following Watson's claim that "the boundaries of moral responsibility are the boundaries of intelligible moral address" ("Responsibility and the Limits of Evil," 258), a wide literature has sprung up that relies on the idea of moral address. See, for instance, Darwall, "Moral Obligation and Accountability"; Macnamara, "Reactive Attitudes as Communicative Entities"; McKenna, "The Limits of Evil and the Role of Moral Address" and *Conversation and Responsibility*; and Telech, "Praise as Moral Address."

tension runs in both directions. On the one hand, if holding others accountable is about imposing burdens on others, then how can it involve an exchange of moral reasons? The most obvious result of imposing burdens—the creation of incentives—seems to have nothing to do with such an exchange. On the other hand, if holding others accountable is about moral address, then what is the point of imposing burdens on them? Why not just try to persuade them?

The concept of reproof provides an answer to both questions. First, holding others accountable through reproof can involve a genuine exchange of moral reasons because the burdens of reproof stand for the moral reasons that have been neglected. Thus, the imposition of burdens plays a key role in reproof's exchange of reasons. Second, the point of imposing burdens in reproof is precisely that they serve a symbolic purpose, allowing us to make a particular kind of moral appeal. Reproof communicates something that cannot easily be expressed in words alone, just like champagne and the tap of a sword. The burdens of reproof make tangible one's bad conduct, as well as one's alienation from the community and the oughts one has violated. In reproof, one can address others in a way that goes beyond verbal expression.

## 3. THE MORAL DIMENSIONS OF REPROOF

### *3.1. Reproof and the Democratic Objection*

Having set out the notion of reproof, we can now address the democratic objection to force. My argument proceeds in two steps. In this section, I argue that in principle, interpreting violent and forceful protest as reproof offers a response to the democratic objection to force. In the next section, I argue that such an interpretation is plausible in actual cases of protest.

The democratic objection to force rests on the claim that violent and forceful protest is incompatible with respect for the outcomes of democratic decision-making. According to the objection, when protestors use force or violence, they attempt to directly bring about their desired political outcome and so force their views on others. But the notion of reproof offers a clear sense in which the use of force or violence need not involve any attempt to force our views on others. When I reprove you for breaking a promise to me, my aim is not to leave you no option but to do as I wish. Instead, I aim to provide you with a reason to keep your promises, albeit a reason that is conveyed through burdensome treatment. It is this reason and not the burdens alone that must bring about a change in your conduct if my reproof is to succeed. If protest can impose burdens in the manner of reproof, then it seems that protestors may use force without attempting to force their views on others and thus do so without acting undemocratically.

We can tease out the idea by comparing this interpretation of forceful and violent protest with that of Ronald Dworkin, who offers perhaps the clearest articulation of the democratic objection. Dworkin describes nonpersuasive civil disobedience as "aim[ing] not to change the majority's mind, but to increase the cost of pursuing the program the majority still favors, in the hope that the majority will find the new cost unacceptably high."[33] Here, Dworkin moves very quickly from the idea that disobedience imposes costs to the idea that it does not aim to change the majority's mind. He therefore describes forceful disobedience in terms that are much closer to sanctioning than reproof. But persuasion is not the only way to change someone's mind, as the idea of reproof makes clear. When we reprove others, we do not provide them with dispassionate arguments. Still, we give them reasons to change their mind. We make a moral appeal, ultimately leaving it up to them whether they will follow through on the reasons we have provided. At no point in this process do we force our views on them, except in the sense that we forcefully tell them how they ought to have behaved. The same is true of reproving protest. As Ted Honderich puts it, "the electorate is restrained or constrained, but in such a way that it is left room for reflection."[34] Democratic decision-making need not be subverted in this process any more than our own will is subverted when we are criticized for bad conduct.

It follows that in cases of burdensome protest like those we began with, we must decide between several different readings. When campus protestors stage encampments, we may read them as using force to force their views on universities, or we may read them as using force to hold universities accountable for their treatment of protestors. When protestors burn down a police station, we may read them as attempting to directly bring about their desired political outcome, or we may read them as using violent means to hold the police or municipal government accountable.[35] Which reading is most plausible depends on the facts on the ground. But it is plausible to think that many such cases look like reproof, especially if the imposition of burdens is designed not to block off alternatives but instead to communicate reasons.

33 Dworkin, "Civil Disobedience and Nuclear Protest," 109. Smith, considering nonviolent uses of force, writes that "a strategy involving ongoing or recurring blockages could only be interpreted as an attempt to raise the cost of a decision in order to prevent its implementation" ("Civil Disobedience and the Public Sphere," 161). This emphasis on cost levying is also found in discussions of coercion in Aitchison, "Coercion, Resistance and the Radical Side of Non-Violent Action," 51; and Livingston, "Nonviolence and the Coercive Turn," 256.

34 Honderich, *Three Essays on Political Violence*, 112. See also Moraro, "Respecting Autonomy Through the Use of Force," 68.

35 The "desired political outcome" here might be a world without police.

Before moving on, there is a caveat. It is compatible with the argument so far to think that there must be an upper limit on the severity of burdens involved in reproof—and therefore also in reproving protest.[36] Suppose that I intend to reprove you but impose burdens on you that are so weighty as to leave you no realistic option but to do as I wish. One might think that in such cases, my actions cease to resemble reproof entirely. One might instead think that my actions resemble reproof in some sense, but they are morally objectionable because they involve forcing you to do as I wish. Alternatively, one might think that such examples show that there is nothing morally objectionable about the imposition of severe burdens so long as they achieve the communicative aims of reproof.[37] There is room for reasonable disagreement here. But minimally it seems plausible that if there are limits to the severity of burdens in reproof, they are unlikely to rule out many of the cases of burdensome protest we are concerned with. However disruptive blockades, vandalism, and arson may be, they do not regularly seem to leave their targets *no choice* but to do as protestors wish.

### *3.2. Categorizing and Justifying Reproof*

So far, I have focused narrowly on the democratic objection to force. But the argument I have offered naturally raises some broader moral questions. Does this argument really imply that keying a car and burning down a police station can be justified simply because these actions can be interpreted as reproof? If so, then it might seem that the democratic objection is not the only relevant moral consideration. If not, then how could my argument succeed as a moral defense of force and violence?[38]

In response, we should be careful to distinguish the categorization of cases as reproof from the justification of reproof. I have insisted on the claim that some protest can be categorized as reproof, which immunizes that protest from the democratic objection. But accepting this claim is compatible with a wide range of views about justification. And although I cannot provide a full account of justification here, it seems reasonable to think that such an account would offer an appealing way of characterizing the justificatory worries raised above. To see what I mean, consider two elements that such a justificatory account is likely to have.

36 I thank an anonymous reviewer for raising this point.

37 On the last point, one might think that imposing such severe burdens on others involves a *risk* of forcing others to do as we wish but that, so long as they act on the reason we give them and not the burdens, they have not been forced to do anything. I thank Alon Harel for bringing this point to my attention.

38 I thank two anonymous referees for raising these issues.

First, there is a range of everyday norms that appear to be internal to our everyday accountability practices. Angela Smith and Miranda Fricker note, for instance, that blame is unfitting when it is directed at the wrong target, when nothing wrong has been done, or when the blame is insufficiently constrained in scope and remit.[39] I take it that the justification of reproof involves a similar set of norms, suitably elaborated. In the case of protest, these norms might suggest that reproving protest is unfitting when it is misdirected, when nothing wrong has been done, or when it is disproportionate. And as Edmund Tweedy Flanigan recently argues, considerations of fittingness may establish a deontic status for protest that is stronger than having a permission and weaker than having a duty.[40] In this way, the justificatory criteria for holding one another accountable interpersonally might suffice to ground a *pro tanto* justification for reproving protest. Second, there are any number of external moral factors that might bear independently on the justification of reproving protest. For example, duties of benevolence, respect for the interests of others, and obedience to the law might all weigh for or against the justification of protest in the final analysis.

These points give us a sense of how a complete theory of reproof might respond to the justificatory worries that motivated this discussion. Let us start with the case of Pat keying the car of Sam, an unfaithful lover. I suspect intuitions will vary about whether this case is justified. But even if one thinks that such actions are unjustified, it seems plausible to say that this is because keying a car is a disproportionate response to infidelity or because such actions are forbidden by external moral considerations. Intuitions are also likely to vary about protestors burning down a police station. But if one thinks that such actions are unjustified, it is likely because they go too far—because they are disproportionate. By the same reasoning, responses are available when protest is directed at the wrong party, or when there is in fact nothing to protest.

It follows that my argument does not require us to suppress our moral worries about controversial cases. To the contrary, it is compatible with different accounts of the internal and external constraints on the justification of reproof, and at least some such accounts can substantiate these worries. The argument is

39 Smith, "On Being and Holding Responsible," 475–76, 478–83; and Fricker, "What's the Point of Blame?" 168–71. I am glossing over much complexity and many differences in presentation, but these differences do not matter for the broad point I am drawing here.

40 Flanigan, "Futile Resistance as Protest," 644–54. Interestingly, Flanigan also defends his account of fittingness by appeal to the idea of holding others accountable. In contrast with my account, however, Flanigan derives substantive criteria of justification by modifying principles of defensive ethics rather than thinking about the internal norms of accountability. Accordingly, I take Flanigan to offer a sympathetic, parallel account of a different form of protest.

designed not to get us to endorse all cases of protest that resemble reproof but instead to help us get clear on the moral stakes of these cases by bracketing a distracting objection. Beyond these points, I leave the issue of justification open.

## 4. INTERPRETING PROTEST AS REPROOF

If we accept the foregoing, then some violent and forceful protest can escape the democratic objection to force on the condition that such protest can be aptly described as reproof. Is this condition satisfied?

Before answering that question, we should note some qualifications. To say the condition is satisfied is not to claim that *all* forceful or violent protest escapes the democratic objection. It would be very surprising if that were true, especially since I have used these terms in an undiscriminating way. Proponents of the democratic objection to force are surely right that some forceful or violent disobedience is undemocratic. To say that the condition is satisfied is also not to claim that *only* violent or forceful disobedience escapes the democratic objection to force. Protest need not be forceful or violent to resemble reproof, even though those are the cases most relevant to the democratic objection. Just as we can reprove others interpersonally through verbal criticism and reprimands, symbolic marches and demonstrations may also count as reproof in some cases. To say that the condition is satisfied is only to say that *some* forceful and violent protest counts as reproof and therefore avoids the democratic objection to force.

With that said, there are two main reasons to think the condition is sometimes satisfied. First, there are strong descriptive grounds for characterizing some disobedience in terms of reproof. On one hand, those who have recently attended a protest have likely heard protestors calling for accountability, chanting "Shame!" and issuing strong moralized demands. For instance, when protestors occupy campus buildings to protest a university's treatment of other protestors, it is not difficult to interpret their actions as resembling the burdensome appeal of reproof.[41] On the other hand, construing the forceful or violent actions of protestors as attempts to force their views on others often overlooks the communicative dimensions of the protest. It is implausible to characterize the George Floyd protests, for instance, as simply attempting to bring about a certain political outcome. It seems essential to note that in burning down a

41 Indeed, some recent work makes an argument going in the other direction, importing the language of protest to characterize blame and not the language of blame to characterize protest. See Smith, "Moral Blame and Moral Protest"; and Talbert, "Moral Competence, Moral Blame, and Protest." Despite the broad resemblance between these views and the view set out here, they draw on Bernard Boxill's view of protest, which assigns protest a very different communicative function. See Boxill, "Self-Respect and Protest."

police station, protestors aimed to *say* something about the injustice of Floyd's murder.[42] Moreover, recalling the earlier caveat, it also does not seem that the burdens imposed in such protests are regularly so substantial as to vitiate an interpretation of the protests as reproof.[43]

Second, to say the condition is satisfied is to make only a parsimonious claim. As we have seen, proponents of the democratic objection to force conceive of nonpersuasive disobedience as using burdens to create incentives for a change in conduct. In other words, proponents of the objection already conceive of protest in terms of an interpersonal moral interaction geared towards achieving a change in conduct. To say that disobedience can resemble reproof is not to substitute an entirely different conceptual framework. Instead, it is to say that if disobedience might sometimes resemble sanctioning, it might also sometimes resemble reproof.

### *4.1. Violence and Force*

I anticipate two main objections to the foregoing argument. First, it might seem that I have inappropriately conflated violent and nonviolent protest throughout this paper. I have argued that protest is sometimes a form of holding others accountable in order to defend the use of force and violence in protest. But we do not accept violence in interpersonal accountability. If you tread on my foot, for instance, it would be outrageous for me to respond by hitting you or spitting on you.[44] So how could the idea of reproof, which is derived from our ideas about interpersonal accountability, capture violent protest?

It is true that we reject violent attempts to hold others accountable in interpersonal contexts. But why? If this is a moral objection—that violent attempts to seek accountability are wrong—then it concerns the justification of reproof and not the main argument of this paper. A stronger version of the objection would be that the concept of reproof excludes violence. Yet this conceptual objection is also problematic. Reproof, as I have described it, involves the communicative imposition of burdens on someone who has acted badly to bring

42 Some recent work in applied philosophy of language bears this out. Michael Randall Barnes, Matthew Chrisman, and Graham Hubb offer analyses of protest as a kind of speech act that bear clear resemblances to my characterization of reproof. See Barnes, "Positive Propaganda and the Pragmatics of Protest"; and Chrisman and Hubbs, "Protest and Speech Act Theory."

43 The Floyd protests might be read as coercively attempting to bring about a world without police. But I think saying this would involve attributing to the protestors a strong political consensus, which seems implausible given the spontaneity and lack of organization of the protests.

44 The example, though not the discussion that follows, is borrowed from Darwall, "Moral Obligation and Accountability," 112.

about a change in conduct. Nothing about this description excludes violence. Now, the objector might respond by arguing that violence is incompatible with the communicative aims of reproof. Echoing Rawls, they might say that violence necessarily issues threats, which is incompatible with the moral appeal of reproof. But as we have seen, violence is a capacious category that can include anything from terrorism to sitting in the middle of a road. Furthermore, violence seems perfectly compatible with a variety of communicative aims. Violence can issue threats, to be sure, but it can also send a message, set an example, or prove a point—consider public executions or, more minimally, corporal punishment.[45] There seems to be no reason to exclude reproof from this list.

*4.2. Justification and Liability*

Second, one might raise a deeper moral objection. As I have noted, it seems plausible to think that reproof more generally and reproving protest more specifically are unjustified when misdirected. It seems to follow that protestors engaged in reproof should take care to impose burdens only on those responsible for the problem they are protesting. Indeed, many defenders of forceful and violent protest argue for similar restrictions.[46] But the burdensome protests we have been concerned with often place substantial burdens on nonresponsible third parties. Extinction Rebellion roadblocks are an inconvenience to the general public, not just to makers of climate policy; and pro-Palestine campus encampments inconvenience a university's student body, not just administrators. Indeed, this may appear to be a general problem for the view I have defended, since it is rarely possible for protestors to directly burden those responsible for the object of their protest.[47]

Conceding this objection would not be devastating for my argument, since protestors do sometimes directly burden responsible parties. It is significant, for instance, that many who protest US Supreme Court decisions choose to do so outside the justices' houses. Still, I think it is possible to say something stronger. To start, the requirement that reproof should be well directed is more

45 I thank Hamish Russell for these examples and discussion of this point.

46 Flanigan writes, "I take expressive acts of protest to be fitting only when they are directed *by* the person who protests *at* the person, or collective, who is responsible for what is protested against" ("Futile Resistance as Protest," 647). Ten-Herng Lai and Chong-Ming Lim similarly argue that burdensome protest has a justificatory advantage when it is directed at those who are responsible for the relevant injustice ("Environmental Activism and the Fairness of Costs Argument for Uncivil Disobedience," 498). See also LaBossiere, "When the Law Is Not One's Own," 328; and Marcou, "Violence, Communication, and Civil Disobedience," 506.

47 I thank an anonymous referee for drawing this objection to my attention.

qualified than it might seem. Even in the context of interpersonal accountability, the burdens we impose on others frequently spill over to nonresponsible parties. If your partner tells a racist joke and is reproved, then the social censure they face might come to diminish your own social status and well-being. But those effects do not seem to make the reproof unjustified. By extension, it seems implausible to claim that any effect on nonresponsible parties makes reproving protest unjustified.

But surely the issue is not just that a protest *inadvertently* affects nonresponsible parties. When Extinction Rebellion protestors block a public road, the inconvenience to the public is not the spillover of burdens directed at the responsible parties. Instead, the issue seems to be that, in many such cases, the imposition of burdens on nonresponsible parties is an essential and intended part of the protest. Still, it seems to me that addressing responsible parties is sometimes possible only *through others*. Suppose a warden mistreats their prisoners through neglect. The prisoners never see the warden, only the warden's officers as they bring food and water. The warden's officers are conscientious and humane, but they are prevented by the warden's orders from speaking with or adequately tending to the prisoners. Under such circumstances, the prisoners may be able to communicate to the warden only through their own treatment of the officers—by throwing food, calling insults, or refusing to eat. In this case, it seems plausible that even if the officers bear the brunt of the burdensome treatment, the warden is still the target of the reproof, properly understood. And given the inaccessibility of the warden, the prisoners' treatment of the officers might well be justified. Generalizing, in cases where a wrongdoer is inaccessible, it may be permissible for reprovers to direct burdensome treatment at intermediaries. Of course, one could deny this claim and insist that reproof is justified only when it imposes burdens primarily on the responsible party. The argument does not stand or fall with this point. But then one must accept that under such circumstances, there is an important sense in which the powerful must remain unaccountable.

If we accept this argument, then reproving protest may be justified even when the imposition of burdens on nonresponsible third parties is an essential and intended part of the protest. In climate protests, the public and the police may bear burdens intended to reprove those responsible for climate policy. In campus protests, disruption in the classroom or on campus may be the best protestors can do to address those responsible for the things they protest.

## 5. Conclusion

In this paper, I have argued that conceiving of protest as a form of holding others accountable allows us to make sense of some cases of burdensome protest and

to defend them against the democratic objection to force. To conclude, I want to tease out the implications of this argument by considering where it fits into the scope of responses to the democratic objection.

Earlier, I contrasted two kinds of accounts: those that accept the democratic objection, with a separate qualification for incidental force and violence; and those that reject the democratic objection and defend a wider range of force and violence. In this paper, I have staked out an intermediate position by accepting the democratic objection while defending a wider range of force and violence. However, it might seem difficult to see how the argument could yield this conclusion. For to accept the democratic objection is to accept an idealizing assumption—that the law is basically democratic—which may seem to rule out circumstances of severe and entrenched injustice. Yet for our account of reproving protest to justify force or violence, it seems, there must be wrongs serious enough to warrant forceful or violent reproof. The task, in other words, seems to be to show that forceful and violent protest can be justified even outside of circumstances of severe injustice. Otherwise, it might seem that I have overstated the disagreement with incidental accounts.[48]

First, I think it is a mistake to characterize the conditions under which the democratic objection obtains as conditions in which there is no severe injustice. Indeed, many influential accounts of civil disobedience take it for granted that nearly just political arrangements may sometimes contain serious injustices.[49] Second, it is also a mistake to assume that reproving protest can be directed only at law or policy. Even if the laws and policies of a state are the result of democratic decision-making, it is entirely possible that police, state officials, corporations, or other bodies may act wrongly in ways that warrant reproof. Third, the argument I have offered is not just a defense of violence but a defense of a wide range of burdensome protest, including blockades, sit-ins, and other nonviolent tactics. Granted that the society in question is basically democratic, wrongs may be committed by state or private entities that warrant that warrant violent or forceful nonviolent responses.

Reflection on accountability, then, furnishes us with concepts that expand the scope of justification for protest. But it also suggests that in democratic politics, as in interpersonal accountability, there is nothing essentially pernicious

48 I thank an anonymous referee for raising this concern.

49 Rawls, for instance, writes that civil disobedience can be justified in a nearly just society as a response to instances of "substantial and clear injustice" ("Definition and Justification of Civil Disobedience," 108).

about being burdened by others. To the contrary, those burdens can be both the surest signs that we have done wrong and our best guides to reform.[50]

*New York University*
*h.krahn@nyu.edu*

## REFERENCES

Aitchison, Guy. "Coercion, Resistance and the Radical Side of Non-Violent Action." *Raisons Politiques* 69, no. 1 (2018): 45–61.

Barnes, Michael Randall. "Positive Propaganda and the Pragmatics of Protest." In *The Movement for Black Lives: Philosophical Perspectives*, edited by Brandon Hogan, Michael Cholbi, Alex Madva, and Benjamin S. Yost. Oxford University Press, 2021.

Bedau, Hugo Adam, ed. *Civil Disobedience in Focus*. Routledge, 1991.

Bennett, Christopher. "The Varieties of Retributive Experience." *Philosophical Quarterly* 52, no. 207 (2002): 145–63.

Boxill, Bernard. "Self-Respect and Protest." *Philosophy and Public Affairs* 6, no. 1 (1976): 58–69.

Brownlee, Kimberley. *Conscience and Conviction: The Case for Civil Disobedience*. Oxford University Press, 2012.

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

———. "Rethinking Civil Disobedience as a Practice of Contestation: Beyond the Liberal Paradigm." *Constellations* 23, no. 1 (2016): 37–45.

Chrisman, Matthew, and Graham Hubbs. "Protest and Speech Act Theory." In *The Routledge Handbook of Social and Political Philosophy of Language*, edited by Justin Khoo and Rachel Katharine Sterken. Routledge, 2021.

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

Coyne, Steve. "The Role of Civility in Political Disobedience." *Philosophy and*

50 This paper was completed while in residence as a Global Fellow in the Hauser Global Law School Program at the New York University School of Law. The paper develops arguments initially formulated in my PhD dissertation, and I am grateful for the help of my supervisor, Sophia Moreau, and committee members Arthur Ripstein and Andrew Franklin-Hall. I am also thankful for comments from Brookes Brown, Alon Harel, Liam Murphy, Hamish Russell, Eric Shoemaker, and two anonymous referees for *JESP*, as well as audiences at the Canadian Philosophical Association, University of Tennessee, Knoxville, University of Toronto, and University of Vienna.

*Public Affairs* 52, no. 2 (2024): 221–50.
Darwall, Stephen. "Moral Obligation and Accountability." In *Oxford Studies in Metaethics*, vol. 2, edited by Russ Shafer-Landau. Oxford University Press, 2007.
Delmas, Candice. *A Duty to Resist: When Disobedience Should Be Uncivil*. Oxford University Press, 2018.
Delmas, Candice, and Kimberley Brownlee. "Civil Disobedience." In *Stanford Encyclopedia of Philosophy* (Fall 2024). https://plato.stanford.edu/archives/fall2024/entries/civil-disobedience.
Duff, R. A. *Punishment, Communication, and Community*. Oxford University Press, 2001.
———. *Trials and Punishments*. Cambridge University Press, 1986.
Dworkin, Ronald. "Civil Disobedience and Nuclear Protest." In *A Matter of Principle*. Harvard University Press, 1985.
Feinberg, Joel. "The Expressive Function of Punishment." *Monist* 49, no. 3 (1965): 397–423.
Flanigan, Edmund Tweedy. "Futile Resistance as Protest." *Mind* 132, no. 527 (2023): 631–58.
Fricker, Miranda. "What's the Point of Blame? A Paradigm Based Explanation." *Nous* 50, no. 1 (2016): 165–83.
Greenawalt, Kent. "Justifying Nonviolent Disobedience." In Bedau, *Civil Disobedience in Focus*.
Honderich, Ted. *Three Essays on Political Violence*. Basil Blackwell, 1976.
King, Martin Luther. "Letter from a Birmingham City Jail." In *Political Philosophy: The Essential Texts*, 3rd ed., edited by Steven M. Cahn. Oxford University Press, 2015.
———. *Where Do We Go from Here: Chaos or Community?* Beacon Press, 1994.
Kling, Jennifer, and Megan Mitchell. "Responsibility and Accountability: Permission for Violent Protest." In *The Philosophy of Protest: Fighting for Justice Without Going to War*. Roman and Littlefield International, 2021.
LaBossiere, Barbara. "When the Law Is Not One's Own: A Case for Violent Civil Disobedience." *Public Affairs Quarterly* 19, no. 4 (2005): 317–30.
Lai, Ten-Herng. "Justifying Uncivil Disobedience." In *Oxford Studies in Political Philosophy*, vol. 5, edited by David Sobel, Peter Vallentyne, and Steven Wall. Oxford University Press, 2019.
Lai, Ten-Herng, and Chong-Ming Lim. "Environmental Activism and the Fairness of Costs Argument for Uncivil Disobedience." *Journal of the American Philosophical Association* 9, no. 3 (2023): 490–509.
Lefkowitz, David. "On a Moral Right to Civil Disobedience." *Ethics* 117, no. 2 (2007): 202–33.

Livingston, Alexander. "Nonviolence and the Coercive Turn." In *Cambridge Companion to Civil Disobedience*, edited by William E. Scheuerman. Cambridge University Press, 2021.

Locke, John. *Second Treatise of Government*. Edited by C. B. Macpherson. Hackett, 1980.

Lyons, David. "Moral Judgment, Historical Reality, and Civil Disobedience." *Philosophy and Public Affairs* 27, no. 1 (1998): 31–49.

Macnamara, Coleen. "Holding Others Responsible." *Philosophical Studies* 152, no. 1 (2011): 81–102.

———. "Reactive Attitudes as Communicative Entities." *Philosophy and Phenomenological Research* 90, no. 3 (2015): 546–69.

———. "Taking Demands Out of Blame." In *Blame: Its Nature and Norms*, edited by Neal A. Tognazzini and D. Justin Coates. Oxford University Press, 2012.

Marcou, Andreas. "Violence, Communication, and Civil Disobedience." *Jurisprudence* 12, no. 4 (2021): 491–511.

Markovits, Daniel. "Democratic Disobedience." *Yale Law Journal* 114, no. 8 (2005): 1898–948.

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

———. "The Limits of Evil and the Role of Moral Address: A Defense of Strawsonian Compatibilism." *Journal of Ethics* 2, no. 2 (1998): 123–42.

Moraro, Piero. "Respecting Autonomy Through the Use of Force: The Case of Civil Disobedience." *Journal of Applied Philosophy* 31, no. 1 (2014): 63–76.

———. "Violent Civil Disobedience and Willingness to Accept Punishment." *Essays in Philosophy* 8, no. 2 (2007): 270–83.

Morreall, John. "The Justifiability of Violent Civil Disobedience." In Bedau, *Civil Disobedience in Focus*.

Rawls, John. "Definition and Justification of Civil Disobedience." In Bedau, *Civil Disobedience in Focus*.

Shoemaker, David. "Attributability, Answerability, and Accountability." *Ethics* 121, no. 3 (2011): 602–32.

Simmons, A. John. "Disobedience, Nonideal Theory, and Historical Illegitimacy." In *Boundaries of Authority*. Oxford University Press, 2016.

Singer, Peter. *Democracy and Disobedience*. Oxford University Press, 1973.

Smart, Brian. "Defining Civil Disobedience." In Bedau, *Civil Disobedience in Focus*.

Smith, Angela. "Moral Blame and Moral Protest." In *Blame: Its Nature and Norms*, edited by D. Justin Coates and Neal A. Tognazzini. Oxford University Press, 2012.

———. "On Being and Holding Responsible." *Journal of Ethics* 11, no. 4 (2007): 465–84.

Smith, William. "Civil Disobedience and the Public Sphere." *Journal of Political Philosophy* 19, no. 2 (2011): 145–66.

Stilz, Anna. *Liberal Loyalty: Freedom, Obligation, and the State*. Princeton University Press, 2009.

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

Talbert, Matthew. "Moral Competence, Moral Blame, and Protest." *Journal of Ethics* 16, no. 1 (2012): 89–109.

Telech, Daniel. "Praise as Moral Address." In *Oxford Studies in Agency and Responsibility*, vol. 7, edited by David Shoemaker. Oxford University Press, 2021.

Waldron, Jeremy. *Law and Disagreement*. Clarendon Press, 1999.

Wallace, R. Jay. *Responsibility and the Moral Sentiments*. Harvard University Press, 1994.

Watson, Gary. "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme." In *Agency and Answerability: Selected Essays*. Oxford University Press, 2004.

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

Welchman, Jennifer. "Is Ecosabotage Civil Disobedience?" *Philosophy and Geography* 4, no. 1 (2001): 97–107.

*Journal of Ethics and Social Philosophy*
VOL. 30, NO. 2 · APRIL 2025
*https://doi.org/10.26556/jesp.v30i2.4363*
© 2025 Authors

# TOUCHING THE GOOD

## SPECIAL RELATIONSHIPS AS CONTACT WITH VALUE

*Adam Lovett and Stefan Riedener*

IMAGINE that your dear neighbor Rachel has had a car accident. It was a foggy night. The rain had formed a thin sheen on the road. She took a sharp corner too fast, and her car spun out of control. She was seriously injured and will be bedridden for weeks. So the household responsibilities have all fallen on her partner, Lynn: Lynn must look after their children, do the housekeeping, manage their business, and so on. You know there are thousands of people in Lynn's situation: single parents, overwhelmed caregivers, struggling businesspeople. Yet Lynn is your neighbor. You have known her for almost your entire life. You have a special relationship with her. So you care much more about Lynn's plight than about the struggles of strangers. You are more moved emotionally by her predicament and will likely do more to actively help her.

We seem to have such special relationships to objects and projects, as well as to people. Consider, for instance, your childhood home. You love the place. You remember it as warm, cozy, welcoming. But the local government wants to build a bypass. They need to knock down your house. The bulldozers threaten many homes that are as valuable as yours. But you have a stronger emotional reaction to its possible destruction than to the destruction of other houses, and you will likely do more to actively protect it. Similarly, imagine you are a mathematician. You have worked for three years on an intricate proof. You think you have finally got it and are about to send it off for publication. But then you discover your mistake: the third lemma on the seventh page is false. Much of your work was in vain. Again, you are not alone in this fate. Many intellectual projects fail. Your failure is not objectively more regrettable than these others. But you are likely to regret it much more and to spend more time trying to repair your invalid proof than the invalid proofs of your colleagues.

In short, we all care more about some things than about other equally valuable things: we care more about the people, objects, and projects with which we have special relationships. This caring has both affective and practical dimensions: we have greater emotional reactions and take more actions in response to these things than to other equally valuable things. Moreover, and crucially, it

seems that this is all as it should be: we are not making a mistake by caring more about our neighbors, our own homes, and our own proofs. Quite the opposite: intuitively, we would make a mistake by ignoring our special relationships. We would go wrong by caring about everything as if from the point of view of the universe. None of us, after all, is the universe. We are individuals, with our own points of view. As we put it, some things are *personally significant* to us. In virtue of our special relationship with them, we should care about them especially.

The aim of this paper is to provide an account of personal significance—of *which* relationships to things make it appropriate for you to care especially about them. Our idea is simple. You can come into contact with a thing's value, and it is this contact with its value—this *touching the good*—that makes it significant to you. We articulate the notion of contact in terms of manifestation: something's value can be manifest in your life, and your life can be manifest in that value; this puts you in contact with it and makes it significant to you. We call this the *contact account of significance* (which will be laid out in section 2). We argue that this account captures both core cases of personal relationships (section 3) and more challenging cases (section 4). Indeed, we suggest that it can also explain our reasons of gratitude and compensation (section 5), as well as how we discount for temporal and modal distance (section 6). Many of our reasons, in short, can be seen as emerging from such contact. So the contact account seems worth taking seriously indeed. But before we turn to this, let us say more about the core phenomenon that we aim to explain.

## 1. THE QUESTION

We start with some stage setting. Intuitively, you care about many kinds of thing: people, places, nations, truths, and so on. But for convenience, we assume that you fundamentally care about facts.[1] More precisely, we assume

1 For expository purposes, it is simplest to assume that our attitudes of caring have just one kind of object. We opt for facts primarily because they are very flexible. It is, for instance, easy to translate care about facts into care about things, and vice versa. One simply says you care about a thing if and only if you care about a fact concerning that thing. Yet we think there is also a more positive reason to focus on facts. There are very many ways in which you could care about, say, Lynn as a bare object. You could want her life to go well for her or could want her to buy your products or to suffer some misfortune. In saying you care about her, we typically have in mind only some of these forms of caring. A straightforward way of putting this is to say you care about certain facts about Lynn but not others—e.g., the fact that her well-being is morally valuable but not the fact that she could increase your wealth or satisfy sadist desires. So to assume we care about facts is often the best way of accurately describing our attitudes. Still, nothing much hinges on the assumption that we care about facts. One can easily reformulate the contact account under the assumption that we fundamentally care about a broader range of entities.

that you care about *value facts*. A value fact is a fact that warrants certain affective and practical responses: certain emotions, desires, intentions, actions, and forms of deliberation. That the Grand Canyon is beautiful, for instance, is a value fact in our sense: it makes it fitting to feel awe and gives us reason to protect the canyon. That the Grand Canyon is dangerous is also a value fact in this sense: it makes it fitting to be slightly afraid and gives us reason to plan carefully before we enter the canyon. In the same sense, that something is admirable, asinine, or funny is a value fact. We say that you *care* about a value fact if you do respond to it with the affective and practical responses it warrants. So for instance, you care about the fact that the canyon is beautiful if you actually feel awe in response to it and try to protect it if you can. You care about the fact that the canyon is dangerous if you are appropriately respectful about entering it and plan accordingly before you do.[2]

Now, you can care more or less about a value fact: you can have a stronger or weaker emotional reaction to it and take more or less action in response to it. And importantly, how much you care can be appropriate or inappropriate. You can care too little about something. Imagine one of your close colleagues has died, but you do not feel anything at all: not an iota of grief tickles your breast. This seems too cold: you should care more about their death. But you can also care too much about something. Imagine your football team loses, and you are utterly overcome with despair: you are laid low for days, consumed with anguish. This seems too warm: you should care less about that defeat. These two cases arise because how much you should care about a fact is partly proportional to the degree of its value. Things can be more or less valuable. And other things equal, you should care more about something the more valuable it is. Thus, your attitude towards things should in part be a function of their value.

Yet how much you should care about a certain fact is not merely a function of that value. It is also a function of your relationship to the fact. In particular, intuitively, your special relationship to a value fact can intensify the weight of your reasons to care about it.[3] This gives rise to our phenomenon. We say that a value fact is *personally significant* to you when you have a special relationship

2 Our notion of caring is similar to Scheffler's notion of valuing. See Scheffler, "Valuing." The main difference is that to value something in Scheffler's sense, you must believe it is "good or worthy" (32). But you can care about something in our sense if you believe it is bad or terrible: you can care about injustice or poverty, say. The notion of caring has been the subject of much other discussion. See, e.g., Seidman, "Valuing and Caring"; and Kubala, "Valuing and Believing Valuable."

3 The idea that personal relationships are intensifiers is emphasized by, e.g., Lord, "Justifying Partiality"; and Löschke, "Relationships as Indirect Intensifiers." It is also suggested in, e.g., Jollimore, *Love's Vision*, 114; Keller, *Partiality*, 136; and Lazar, "The Justification of Associative Duties," 51.

towards it, and this makes it appropriate for you to care more about it than it would otherwise be. Our three cases exemplify this phenomenon. You should care especially about the fact that your neighbor is struggling, that your house might be destroyed, or that your work has failed. You should care more about these facts than about other, objectively similar value facts, due to your relationship with them. But there are many more such cases: you should care more about the cuteness of your child than about the cuteness of some random child somewhere far away; more about the beauty of a sunset today than about that of a similar sunset a year ago; more about the injustice of racism in your town today than about its injustice in third-century Rome. In each case, intuitively, you have a special relationship with the first fact. And that is why you ought to care about it especially. It is why, in our parlance, it is personally significant to you. Our question is: What makes a fact personally significant to you in this sense?

To be clear, the target of our inquiry does not cover all cases in which you have some reason to care about one thing more than another equally valuable thing. Imagine a billionaire says that they will give you a million dollars if you care especially about their favorite book. This arguably gives you some reason to care about that book more than other equally good books. We do not aim to capture this phenomenon. In such a case, your reason to care especially about the book is not explained by any substantive special relationship you have with it. You do not (yet) have any such relationship with the book. Your reason to care about it is grounded in the billionaire's offer alone. We are interested in cases in which your reason to care about something especially is grounded in such a special relationship. Here we get a grip on the pretheoretical notion of a special relationship by seeing it at work in our paradigmatic examples—your relationship to your neighbor, your childhood home, your own proof. The target of our inquiry is that sometimes such relationships make it appropriate to care especially about something. When we aim to explain *personal significance*, we aim to explain just this phenomenon.

Let us stress something. As we have glossed the matter, it is not just people that can be significant to you: objects and projects and indeed any kind of value fact can be significant to you. Personal significance is a very general phenomenon. That is not to deny that there are important differences here. For instance, your reasons to care about different things will often be of different kinds. When you have a special relationship to a person, it is typically *morally* inappropriate not to care about them. You do something morally wrong if you are indifferent towards their value. When you have a special relationship to an object or project, moral considerations are less often at issue. Indifference towards the value of your house or proof will typically be an insensitivity to aesthetic, historical, or prudential value. Still, a vast range of things can be

personally significant. So we think that, other things equal, it is preferable to have an account of personal significance that respects the relevant differences here, at the same time as being general and unified.[4]

Many accounts of significance have been proposed. Yet we find none of them convincing. Our reasons are for the most part familiar. But it is worth rehearsing them briefly. Let us start with subjectivism. According to subjectivism, what makes something personally significant for you is your subjective desires, commitments, or concerns for that thing. You have special reasons with regard to your proof, say, just because you care about it. Bernard Williams is sometimes interpreted as defending such an account, in the form of a "project view."[5] We think such subjectivism is too subjectivist. It implies that if you just do not care about your neighbor, house, or proof, you do not have any special reasons regarding them. And if out of a psychological idiosyncrasy, you care especially about people with the same number of hairs or the same skin color as you, you have special reasons vis-à-vis these people. But this, we think, is wrong. If you have known your neighbor for decades, you would just be wrong to treat her like a complete stranger. And your peculiar psychological attitudes do not give you reasons to care especially about the people who share your hair number or skin color. Personal significance is at least partly an objective matter.[6]

A natural way to introduce objectivity is through the value of the relevant relation. That is what relationship views do. On these views, what makes something personally significant for you is the fact that you have a noninstrumentally valuable relationship to it. So you have special reasons with respect to your childhood house, say, because your relationship to it is noninstrumentally

4 For a defense of the importance of unity in the context of moral theories, see Brink, *Moral Realism and the Foundations of Ethics,* 249–52. For a more general discussion of it, see Keas, "Systematizing the Theoretical Virtues," 2775–80.

5 See Williams, "Persons, Character, and Morality." For this take on Williams, see Keller, *Partiality*, 31–35. Williams's broader project is to give an account of reasons in general: to say that *all* our reasons are somehow grounded in subjective states (see esp. "Internal and External Reasons"). One need not be such a general subjectivist to endorse subjectivism about significance. One might think there are objective values but that our subjective attitudes give us reasons to care especially about some of them rather than others. Still, general subjectivism is compatible with how we characterize the phenomenon of personal significance. For instance, even as a general subjectivist, Williams can (and, we think, does) give a specific account of the *particular* reasons that arise from special relationships in our sense. We focus on just this specific aspect of his overall view. For some similar other versions of this account of personal significance, see Frankfurt, *The Reasons of Love*; Stroud, "Permissible Partiality, Projects, and Plural Agency"; and Drake, "Love, Reasons, and Desire."

6 A similar worry about subjectivism is articulated in, e.g., Scheffler, "Projects, Relationships, and Reasons"; and Keller, *Partiality*, ch. 2. For more arguments, see, e.g., Jeske, "Friendship, Virtue, and Impartiality" and *Rationality and Moral Theory*.

good. Such a view is defended, for example, by Samuel Scheffler.[7] Relationship views are adequately objectivist. Your relation to your neighbor (or the people sharing your hair number or skin color) may have (or lack) value and thus give (or not give) you special reasons—whether you want it or not. Yet relationship views face a familiar and, we think, serious focus problem. They locate the relevant value in the *relationship* you have to something. But intuitively, it is in the *thing itself*. If you treat Lynn like a complete stranger, the core problem is not that you fail to value the relationship you have to her (your neighborhood or your friendship); it is that you fail to value *her*.[8]

This has led many people to defend individualist views. According to these views, what makes something personally significant for you are the individual properties or value of the thing itself. For instance, ultimately, Lynn is personally significant to you because she is so benevolent and kind. This kind of view is defended most prominently by Simon Keller.[9] It avoids the focus problem as it locates the relevant value in the object of significance itself. However, it faces a challenge. It needs to explain *which* things we ought to care about especially and *why* we ought to do so—given that many other things are objectively just as valuable. There are many people as benevolent and kind as Lynn. So why is it appropriate for you to care especially about her? Or why would it not be appropriate to care especially about an otherwise similar woman you have never met and know nothing about?[10]

A natural line to take at this juncture is to emphasize something like personal acquaintance. We may call this the acquaintance view. On this view, what makes something personally significant for you is that you are acquainted with its individual properties or value. Here acquaintance is the kind of relationship that you have to your childhood home, for example. You have experienced it, have a special understanding and knowledge of it. The proposal is that something is significant to you because of such acquaintance. Such an idea is

7 See Scheffler, "Relationships and Responsibilities," *Boundaries and Allegiances*, "Morality and Reasonable Partiality," and "Membership and Political Obligation." Related views are suggested by, e.g., Raz, "Liberating Duties"; Seglow, *Defending Associative Duties*; and Lazar, "The Justification of Associative Duties."

8 See Keller, *Partiality*, 62–64. In our official language of value facts, the problem here is that what we care about when we care about someone is not a relational fact (i.e., how they are related to us) but facts singularly about them: their well-being, their mental states, their life.

9 See Keller, *Partiality*. Related ideas are defended by, e.g., Velleman, "Love as a Moral Emotion"; Lord, "Justifying Partiality"; and Naar, "Subject-Relative Reasons for Love."

10 Keller himself mentions this worry (*Partiality*, ch. 5). In response, he accepts a form of particularism (150–52). Many people have found this view unsatisfyingly nonexplanatory. See, e.g., Olson, "Review," 624–26.

suggested, for example, by Kieran Setiya.[11] *Prima facie*, it helps meet the challenge set out for the individualist view. We are, after all, acquainted with only certain things. So perhaps acquaintance views can explain why we should care about some things more than others. But such views are also unsatisfactory. The problem is that, as we understand the relation of acquaintance, it is too passive. Acquaintance involves the world impinging on you rather than you impinging on the world. Yet sometimes how you impinge on the world matters to what you should care about. Consider your mathematical proof. In principle, you might be just as acquainted with a colleague's attempt to prove something as with your own. Your colleague might keep you exhaustively updated on their progress: on their blind alleys, their small successes, their ultimate failure. You might even understand more about their proof than about yours: you might have a keener awareness of why their strategy did not work or of what this means for their overall research project, or you might have forgotten some details of your own work. Still, you have a connection to your own proof that you cannot have to your colleague's: you *produced* it. You do not merely understand it; you made it. This is (at least partly) why your own proof is typically more significant to you. Acquaintance does not capture this active dimension of personal significance.

One might think that the resonance view, as defended by Niko Kolodny, captures this active dimension.[12] Kolodny thinks that our attitudes should be, in a special way, coherent. For instance, suppose the discrete encounters you have had with Lynn called for sympathy. Then your overall attitude to her should also be a kind of sympathy—but a sympathy that is open ended, involves a kind of commitment for the future, and thus reflects the fact that you have had a prolonged history with her.[13] Or again, suppose that you and Lynn were both active in opposing the plans to build a power plant in the region. Then you may have reason to step in for her while she is occupied with Rachel's accident—and not just instrumentally to secure your project but also out of solidarity with her, as a reflection of the fact that you were both in this project together.[14] Kolodny sees these cases as instances of a more general phenomenon, which he calls *resonance*:

11 See Setiya, "Other People." Similar thoughts are explored in, e.g., Jollimore, *Love's Vision*; Lewis, "The Aesthetics of Coming to Know Someone"; and Kirwin, "Value Realism and Idiosyncrasy."

12 See Kolodny, "Which Relationships Justify Partiality? The Case of Parents and Children" and "Which Relationships Justify Partiality? General Considerations and Problem Cases."

13 Kolodny, "Which Relationships Justify Partiality? The Case of Parents and Children," 51.

14 Kolodny, "Which Relationships Justify Partiality? The Case of Parents and Children," 52.

> *Resonance*: One has reason to respond to *X* in a way that is similar to the way that one has reason to respond to its counterpart in another dimension of importance, but that reflects the distinctive importance of the dimension to which *X* belongs.[15]

In our examples, you have reason to respond to Lynn's situation in a way that is similar to how you would respond to a stranger suffering or missing in the policy rally—but that reflects the fact that you have had a history of positive encounters ("resonance of histories of encounter") and shared activism ("resonance of common personal situations") with Lynn. Perhaps this can capture both the passive and the active dimensions of significance: resonance plays out in many ways. But this, we think, also indicates a bug: the resonance view is not really that unified. Kolodny understands resonance not as an underlying law that grounds the personal significance of various facts. Rather, it is "merely a description of an abstract structure shared by principles" that do the actual explanatory work—like resonance of histories of encounter or common personal situations.[16] According to Kolodny, when we explain why you ought to be partial towards your neighbors, your children, or your parents, we still need to invoke a host of substantively different actual principles—appealing sometimes to the normative import of responsibility, to that of having an aim, and so on. Resonance provides a unified *form* of explanation, but it does not point us to any single explanans. In this sense, we might say the unity provided by resonance is shallow. Other things equal, a deeper unity would be preferable.

So these are our reasons for proposing a novel view. In sum, we think no existing account of personal significance is fully satisfactory. We want an account that respects our intuitive judgements while being partly objectivist. Yet we also want an account that says that in special relationships we respond to the value of that to which we are related. And we want an account that unifies the different cases of personal significance while respecting both active and passive kinds of significance. Our aim in this paper is to offer such an account. We think the idea of manifestation is crucial for that. So let us turn to this now.

## 2. THE ANSWER

We start from a simple intuition. The unifying phenomenon at play in our examples is that you are in a kind of *contact* with the value of these things—the value of your former home or of your mathematical proof. More precisely, you stand in passive contact with the value of that house: you have been affected

15 Kolodny, "Which Relationships Justify Partiality? The Case of Parents and Children," 47.

16 Kolodny, "Which Relationships Justify Partiality? The Case of Parents and Children," 47.

by it. You stand in active contact with the value of your work: you have affected it yourself. You do not stand in such contact with the values of other buildings and other intellectual endeavors. And that is why you ought to care especially about your home and your proof. More generally, there is a vast universe of value facts out there, but you are in close active or passive contact with just a fraction of them. And a value fact is significant to you only insofar as you are in such contact with it.

How ought we to understand this contact? One simple idea is that contact is a causal notion: you are in contact with a value fact $p$ to the extent that $p$ has a causal impact on you or that you have a causal impact on $p$. Your house had a causal impact on you, and you had a causal impact on your proof: thus, you ought to care especially about them. However, this view is overinclusive. Consider Rachel's accident again. Suppose two people administered first aid after that crash, fell in love on the spot, and later started a family. Imagine that one of their children became a painter of beautiful paintings. Rachel was a chief causal contributor to the beauty of these artworks: if it were not for her, they would never have been painted at all. But the fact that these paintings are beautiful does not seem significant to her: she does not seem to have special reasons to appreciate them aesthetically. They were an all too fortuitous consequence of her actions. More generally, mere causal connections often seem too contingent or incidental. A more internal or nonaccidental connection to something's value underpins personal significance.

We propose therefore that contact should be understood in terms of *manifestation*.[17] Manifestation is a nonaccidental causal connection. At its core, it can be understood in terms of dispositions. Consider the fragility of a vase. This is the disposition to shatter when dropped. We say that its being dropped is the *stimulus condition* of the disposition, and its shattering is its *manifestation condition*. If $D$ is a disposition with stimulus condition $S$ and manifestation condition $M$, we say that $M$ *actualizes* $D$ when $M$ occurs because $D$ and $S$ obtain. In this sense, the vase's shattering actualizes its fragility when it shatters because it is fragile, and it is dropped.[18] At a rough approximation, manifestation is just actualization of dispositions. When you drop a vase, its shattering manifests its fragility. Now perhaps since you have broken that vase, you need to refund the owner for their loss and thus get into financial trouble. Your troubles are caused by the fragility of that vase but do not manifest it. Fragility is not the disposition to cause financial troubles. It is the disposition to shatter. So only the shattering

17 We have employed the notion of manifestation already in Lovett and Riedener, "Commonsense Morality and Contact with Value" and "The Good Life as the Life in Touch with the Good." In the following exposition, we draw on this work.

18 For an overview on the metaphysics of dispositions, see Choi and Fara, "Dispositions."

manifests the fragility. People's dispositions can also be manifest. Leonardo da Vinci was disposed to produce beautiful artworks. The beauty of *The Last Supper* manifests this disposition. Joseph Stalin was disposed to mercilessly vanquish his enemies. Leon Trotsky's death manifested this. Manifestation, in all these cases, is less accidental than brute causal connections. Thus, at a rough approximation, we say you are in contact with a value when your dispositions are actualized in it, or its dispositions are actualized in you.

But that is only a rough approximation. The problem with it is that, in this sense of 'manifestation', only dispositions are manifest. But plausibly, manifestation is a little more than mere actualization of dispositions. To see this, consider looking at a beautiful painting. One might think the beauty of the painting is not itself the disposition to cause aesthetic experiences in people who look at it. The beauty and this disposition are not identical. Still, those aesthetic experiences do, in a perfectly natural sense, manifest the painting's beauty. Similarly, imagine you write a book of beautiful poems, yet your work is unjustly reviled by the critics. One might think your talent is not itself a disposition to cause unjust criticism. Talent might be a disposition to cause adulation, but not unjust revilement. Yet again, in a perfectly natural sense, the fact that this criticism is unjust is a manifestation of your lyrical prowess.

To capture this, we use the notion of ground. Think about the connection between crimson and red, between the members of a set and the set as a whole, or between the parts of a table and the table. These all exemplify a distinctive kind of noncausal explanation—grounding.[19] Similarly, the connection between a painting's beauty and its disposition to cause aesthetic experiences is a grounding relation. The fact that your poems are good grounds the fact that criticism of them is unjust. Roughly, we want to say that manifestation is indifferent to such grounding connections. More precisely, let us say that $p$ is *ground-theoretically connected* to $q$ if and only if $p$ grounds $q$ or $q$ grounds $p$.[20] We say that $q$ *manifests* $p$ if and only if $q$ or something ground-theoretically connected to $q$ actualizes $p$ or something ground-theoretically connected to $p$.[21] This definition picks out a class of nonaccidental connections. It precisely

19 For more on grounding in general, see Rosen, "Metaphysical Dependence"; and Fine, "Guide to Ground." One might worry that these cases are too disunified to pick out any one relationship. But we think they all pick out a kind of noncausal explanation, and one can say general things about the formal features of this form of explanation (e.g., its transitivity and asymmetry) and even how it connects to modality. That is unity enough for the work we want grounding to do in this paper.

20 The relevant notion of ground is strict partial ground. See, e.g., Fine, "Guide to Ground."

21 These definitions are originally from Lovett and Riedener, "Commonsense Morality and Contact with Value," 414.

articulates our conception of the kind of contact with value relevant to personal significance. A little diagram (figure 1) helps to illustrate this notion.[22]

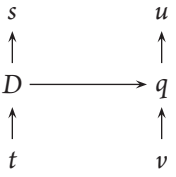


FIGURE 1

The left-to-right arrow stands for the actualization relation while the upward arrows stand for the grounding relation. So in this diagram, $q$ actualizes disposition $D$, while $t$ grounds $D$, $D$ grounds $s$, $v$ grounds $q$, and $q$ grounds $u$. Let us see how this applies to our cases. Imagine that you write a book of beautiful but unjustly denigrated poems. Say that your lyrical talent is a disposition $(D)$ to write beautiful poems (rather than to cause unjust revilement), and the beauty of these poems $(q)$ actualizes that disposition. And say the fact that your poems are beautiful grounds the fact that that criticism is unjust $(u)$. Then, we say that the injustice of that criticism is a manifestation of your lyrical talent. Similarly, imagine that you look at a beautiful painting and have an aesthetic experience. Say that the painting has a disposition $(D)$ to cause aesthetic experiences when people look at it. Your aesthetic experience $(q)$ actualizes this disposition. Suppose that this disposition is grounded in the painting's beauty $(t)$. Then, we say, your experience manifests the beauty.

Again, we suggest that manifestation in this sense constitutes the relevant contact with value. For you to be in contact with a value fact $p$ is for some fact in your current life to be connected with $p$ through a manifestation relation of this form. More precisely, we propose the

> *Contact Account of Significance*: A value fact $p$ is significant to you to the extent that (1) $p$ is manifest in your current life or (2) your current life is manifest in $p$.

The first clause here picks out the passive aspect of contact. In this way, when you are elated by the beauty of a painting, that puts you in contact with its beauty: your elation is a manifestation of it. The second clause picks out its active aspect. In this way, when you write a beautiful poem, that puts you in contact with its beauty: your aesthetic sensitivity is manifest in it. It is such

22 We use the same diagram in Lovett and Riedener, "Commonsense Morality and Contact with Value," 414.

contact that gives you reason to care especially about this beauty. The contact intensifies the strength of your reasons to care about the beauty. The more closely you are in contact with a value fact, the more you should care about it. This is the heart of our view.

Let us make two more clarifications. First, what do we mean by 'your life'? We are thinking of a life as a collection of facts. It includes all facts about what you ever do, feel, see, believe, or desire. Thus, if you once wanted to complete a marathon, think running is a good test of character, or remember your marathons with fondness, those facts are part of your life. Your life also includes all the dispositions you have ever had. So if you are disposed to be exhausted for days after running a marathon, that is part of your life as well. This is not a fully general account of what is a part of your life, but it gives us a reasonably good intuitive fix on it. Note, however, that according to the contact account, it is only your *current* life that matters to significance: you should care about something insofar as it is manifest in your life right now, or your life right now is manifest in it. So what matters is only your present beliefs, memories, abilities, and so on. Past feelings, desires, or dispositions are irrelevant. Imagine, say, that you visited Massachusetts as a small child. The state might have played a big role in your infancy. But suppose nothing about it is manifest in your current life, or vice versa: you have not retained, say, fond memories of clam chowder or bad baseball. Then you should not care especially about Massachusetts. It is only the present that counts.

Second, what sets the extent to which a value is manifest in your life, and vice versa? We think this is (perhaps among other things) a matter of how *central* the manifestation relata are to your life and to a value fact. On the one hand, some things are more central to your life than other things. Your love for your children is more central to your life than your aversion to toads. On the other hand, some things are more central to values than other things. The Sistine Chapel's beauty is more central to the Vatican's magnificence than is that of the papal altar. This intuitive notion of centrality can be interpreted ground-theoretically: $x$ is more central to $y$ than $z$ insofar as it grounds $y$ to a greater degree. The overall shape of your life is grounded in, say, both your love for your partner and your aversion to toads. But the former grounds it to a greater degree and so is more central to it.[23] You are more intimately connected to a value insofar as things more central to your life are manifest in things more central to the value, or vice versa.

23 The idea that grounding comes in degrees has not yet been explored in the literature. But it seems very intuitive. Also, it is quotidian to analogize grounding to causation (see, e.g., Fine, "Guide to Ground"), and causation clearly comes in degrees. So we see this lack of exploration as a shortcoming in the literature rather than in the idea.

So that is the contact account. One might at this point wonder *why* contact with value makes a thing personally significant to you. Why should it have such normative import? The force of this question is defused by seeing the wider import of contact with value. In other work, we have tried to make that wider import clear. In our paper "Commonsense Morality and Contact with Value," we argued that contact with value can explain many features of commonsense morality. It can explain, for example, our obligations to keep our promises and not to harm others. In "The Good Life as the Life in Touch with the Good," we independently argued that contact with value can furnish us with a theory of personal well-being. The idea is that the good life is the life in contact with the good. The contact account of significance corresponds to the reactive contact principle advanced in "Commonsense Morality and Contact with Value" and uses the same core ideology as all of this work. If these other claims are true, contact with value has a broad application across moral philosophy. One need not of course accept these other views to accept the contact account of significance. But contact with value constitutes a unified and general explanation of various ethical domains. And this makes it much less surprising that it would help us explain the phenomenon of personal significance specifically.

There is still a general question about whether contact principles have any deeper explanation. We need not commit to an answer to that question here. But our own view is that normative inquiry has to stop somewhere. A set of very broadly explanatory principles is a good stopping point. So we are inclined to take the contact account to state a fundamental fact, a place in the normative universe where the spade is turned. We are not opposed to a more fundamental explanation of the account, but we personally do not know of any—and think that none is necessary.

Still, we have not yet provided any evidence for the contact account. We turn to that now. Our argument for the contact account is abductive: it rests on how well the account explains cases. So let us now see how the account illuminates our initial examples.

### 3. PEOPLE, OBJECTS, AND PROJECTS

Consider the case of your neighbor Lynn. It is bad for her that she is overwhelmed by the situation. Part of what grounds this badness, say, is that she is such a compassionate, selfless, and caring person. She has a strong desire that the people around her fare well. If she did not have this desire—if she were more insouciant about Rachel's bedriddenness or about the sorrows and woes of their children—her situation would not be as bad for her. Lynn's concern for others can be understood in terms of dispositions: she is strongly disposed

to empathize with other people and help them when she sees they are in need. And these dispositions are manifest in your current life. Your business is afloat today because of the money she lent you. You have got through many a difficult night and are now more serene than you have been due to her emotional support. You stand in a relationship of trust and mutual dependability with her precisely because she is so generous, warm, and considerate. So the fact that Lynn's troubles are bad for her is manifest in your current life. Hence, it is significant to you: you should care about it especially. So in virtue of your relationship to Lynn, you should empathize especially with her struggles and be extra motivated to support her.

But Lynn's general benevolent nature might be only a part of the overall story. Perhaps some more specific dispositions of hers are manifest in your life too. Consider her desire to improve the welfare of her children, say—her disposition to help them when she sees they are in need, to ask others for support in this, get emotional about their welfare when it is brought up in conversation, etc. These more specific dispositions too partly ground the badness of her situation. And perhaps they are also manifest in your life. Perhaps you are currently lacking your cake pan, as Lynn borrowed it for her daughter's birthday. Perhaps you have many memories of looking after these children when Lynn asked you to do so. Or you have a vivid image of her as a concerned mother because she has often been emotional about it. In addition, perhaps you stand in some active contact with her predicament as well. Perhaps part of what grounds the badness of her situation is that Lynn's family are currently a little short of money. And perhaps this fact manifests your readiness to accept favors from others: you accepted their loan, even though they were low on funds themselves. The general point is simple. The more of your life you have shared with Lynn and the more of her life she has shared with you, the more you are in contact with the value of her life—or the badness of her current plight. Your interactions constitute a web of mutual manifestation relations. And this gives you especially weighty reasons to care about Lynn's troubles.

A similar story also goes for other interpersonal relationships. Consider relationships between children and parents. Usually, parents' valuable properties are manifest in valuable parts of their children's lives and vice versa. A parent's concern for their offspring is manifest in that child's flourishing. The child's flourishing is manifest in their parent's joy. Their need for care is manifest in the parent's loving attention, and the loving attention is manifest in the child's emotional development.[24] Something similar is true of good friendships. Friends' valuable properties are manifest in valuable parts of their lives. Your

24 For some related points, see Brighouse and Swift, "Legitimate Parental Partiality," 53–54.

friend takes joy from your conversational acumen. You get pleasure from their sharp wit. They profit from your stout dependability. You benefit from their empathy. You have disclosed your inner life to them, and they have opened themselves up to you. More generally, when you have a special relationship with someone, you are especially in touch with their value, and they are especially in touch with yours. That is why you should care more about your nearest and dearest than about people more distant from you.

Now let us see how this applies to objects and projects. Think about your childhood home. The fact that this house might be demolished is bad. Part of what grounds this badness is that the house had certain dispositions: it was disposed to make people feel at home in it, to make them love the place or have fond memories of it. Your love for that house and your memories of growing up in it manifest these dispositions. In contrast, none of your memories or feelings manifest the value of houses on the other side of the planet. So the threat to your house is especially significant to you. You should be especially moved by it and perhaps try especially hard to stop your home's demolition. Here too the point generalizes. The value of your dearest things—your treasured bicycle, that magnificent island where you spent many a summer, the tradition of klezmer music of which you are so fond—are manifest in your life. You have extra reason to care about them.

Similar points apply to your failed mathematical proof. Part of what grounds the regretability of this failure is that, apart from the lapse in that lemma, the work was excellent. This excellence manifests your intelligence, creativity, and patience: you are disposed to do excellent work in mathematics, and your proof manifests that. In contrast, the proofs of your colleagues manifest their creativity rather than yours. So the waste of your work is especially significant to you. You should regret it more than you regret other failures and perhaps try especially hard to save what can still be saved. And again, the point seems to generalize. You are manifest in the value of your projects—your own Klezmer band, your five-person family, or the shared striving for justice in your country to which you contribute a tiny bit. So you are in close contact with the value of all of these projects and have extra reason to care about them.

These different domains of personal significance do differ somewhat. Relationships with people, for example, give rise to moral reasons, whereas those with objects usually do not. Fortunately, the contact account can also respect these differences. The key point here is that different domains are associated with different kinds of value. People have moral value: their lives have moral import, their autonomy has moral weight, and their virtues have moral worth. But objects and projects are less thoroughly infused with morality. Your projects typically have prudential value: the success of your proof would be prudentially

good for you. Some objects have aesthetic value: your home might have been elegant, graceful, or stunning. Moral value gives rise to moral reasons, but other sorts of value generate nonmoral reasons. Thus, contact with the value of people generally intensifies your moral reasons while that with the value of objects and projects is more likely to intensify your prudential or aesthetic reasons.

So the contact account seems to explain our initial cases well. In fact, it seems to meet all the criteria we have sketched in section 1. Note that whether or not you are in a manifestation relation with something is not entirely subjective. You can be in (or lack) contact with something whether you want it or not. Also, the relevant value that is doing the normative work is that of the thing itself, not that of your relationship to it. These things are significant for you because you are in contact with *their value*, not somehow because it is valuable for you to have such contact. Still, the account is thoroughly unified. Be it active or passive or relating to people, objects, or projects, the relevant relation is always the same. These are sufficient reasons alone to think the contact account is a very good account of significance: that contact with a value intensifies the weight of the reasons this value grounds. But the main argument for the account lies in its broader explanatory power.

## 4. BEYOND THE CORE CASES

Plausibly, the more explanatorily powerful a view is, the more seriously we should take it. So we now turn to how the contact account can illuminate some phenomena that go beyond our initial cases of personal significance.

We start with a case of chosen attachment. Above, we have claimed that personal significance is not entirely subjective. *Inter alia*, that means you cannot always just choose which things are significant to you. Still, sometimes your choices matter. Suppose you become attached to your local football team. You come to deeply value its victory, to admire its sporting prowess. This, it seems, can make it appropriate to care more about this team than about its rivals. This phenomenon is common: often we choose what to care about, and the ensuing care is perfectly appropriate. The contact account straightforwardly explains this. Typically, a characteristic manifestation of value is a certain kind of valuing. Your team's sporting prowess is a disposition that manifests itself in people's admiration. When you admire your team, you are in contact with that value. Likewise, when you take joy in your team's victory, you are in contact with the goodness of that victory. Positively appraising a good typically manifests that good. The contact account can then explain quite generally how we are able to choose our attachments. The crucial point is simply that contact with value need not *always* be objective: it can consist in your own attitudes towards that value.

Let us consider a second case. Suppose you have just had a daughter. You should care immensely about her. Indeed, you should care more about her than you care about even your old friends. Yet your friend's good qualities might seem more manifest in your life than that of your newborn child. Your daughter has been alive for only a few hours: she has not, one might think, had time to manifest her virtues in your life. Nonetheless, the contact account can explain why you should care more about your newborn daughter than about your longtime friends. For a start, you can choose your attachments. When you value your child's innocence or care about her vulnerability, that puts you in contact with her innocence and vulnerability. Valuing your child puts you in contact with its value. Additionally, your child's value might not yet be much manifest in your life, but you are manifest in the child's life. You created the child and helped sustain her through pregnancy. This puts you in contact with the valuable features of your newborn, and that is why you should care about her especially. Contact with value is not only passive but also active. Such active contact with value, we suspect, explains the significance of biological descent.

Let us turn to a third phenomenon. We have focused on cases of positive partiality—when you have reason to feel positively towards someone or something because of your special relationship with it. But there are also cases of negative partiality.[25] Imagine you have an enemy who has undermined you at every turn. They have cruelly trashed your work, wantonly broken up your relationships, and maliciously frustrated your goals. You need care far less about your enemy's flourishing than that of a stranger. Indeed, perhaps it is appropriate to hope your enemy's life goes badly. Again, the contact account explains this elegantly. The fact that someone is cruel and malicious is a negative desert base: it makes it good for that person's life to go badly or at least undercuts the goodness of it going well. When someone is cruel to you, you are in contact with their cruelty. Hence, you have especially weighty reason to want their life to go badly or not to want it to go well. You have reason to want your enemy's life to go badly or to temper your desire that it go well because you are in contact with their vices. More generally, negative partiality arises when you are in contact with someone's bad features. These features warrant negative rather than positive responses.

Consider a fourth case. The contact account, we have argued, can explain both positive and negative partiality. But some relations give rise to no duties of partiality at all. Think about hair number or skin color. The fact that you have the same number of hairs as a stranger does not give you any reason to care

25 For illuminating recent discussions of negative partiality, see Brandt, "Negative Partiality"; and Lange and Brandt, "Partiality, Asymmetries, and Morality's Harmonious Propensity."

about them especially. The fact that you have the same skin color as someone else gives you no reason to be partial to them. The contact account explains this straightforwardly. Sharing the number of hairs with someone is not a way to be in contact with their good qualities. It does not make their value manifest in your life or vice versa. And so according to the contact account, shared hair number does not underpin duties of partiality. This point generalizes to all cases in which sharing a feature with someone does not mean you should care about them especially.

We will end with a fifth case. Imagine you spend five minutes speaking to someone about their life. You have a closer relationship to them than you do to a perfect stranger. Yet such a passing acquaintance does not seem enough to justify partiality in very important decisions. If you can later save either the life of your passing acquaintance or that of the perfect stranger, your closer relationship with the former does not mean you ought to favor them. The contact account can explain this. The explanation invokes incommensurability. Lives are typically incommensurable with respect to value. That means our reason to save one stranger is typically not weightier than, not less weighty than, nor exactly equally as weighty as our reason to save another stranger. These reasons are incommensurable in weight. A small addition in weight to the reasons to save one stranger's life usually does not disturb such incommensurability.[26] But a passing acquaintance with someone establishes only a very mild form of contact, and so it only very mildly intensifies the reason you have to save their life. Thus, these reasons remain no more weighty than your reason to save a perfect stranger's life. Passing acquaintance is not generally enough to make a difference to life-and-death decisions.[27]

The contact account, then, provides an elegant, unified explanation of precisely those cases we want an account of personal significance to explain. This explanatory power speaks strongly in the account's favor.

## 5. GRATITUDE AND COMPENSATION

We have seen how the contact account can explain some clear cases of personal significance. We now turn to some novel cases. These cases are not, on the

26 For a classic discussion of this point about "small improvements," see Chang, "Introduction."

27 Perhaps there is something else at work in such decisions too. You might have some positive reasons of fairness to exclude considerations of partiality from your deliberations. Perhaps it is unfair to let your relationship with someone affect whether you save their life. You have reasons of fairness to be impartial. If so, this would also help explain why you should not let your passing acquaintance with someone affect whether you save their life. We will let readers decide whether they prefer this explanation or the one in the text.

surface, standard examples of special relationships. But one of the virtues of the contact account is that it lets us understand such cases in relational terms. The contact account lets us expand the circle of partiality, so to speak: it lets us understand more phenomena as examples of personal significance than we could without it. In this section, we explain how gratitude and compensation can be understood as kinds of partiality. The ability of the contact account to facilitate such an understanding is part of the abductive argument for it.

We begin with gratitude. When someone benevolently benefits you, you should be grateful. This means in part that you should care especially about them doing well: you should be extra moved if they are doing badly, say, and extra motivated to help them out when you can. Here is how the contact account can explain this. Your benefactor's life is valuable: it is good if they are doing well and bad if they are doing badly. This value fact generally gives rise to reasons: it is appropriate for us all to hope that your benefactor does well and to help them out if we can. Now, part of why your benefactor's life is valuable in this way is that they are virtuous: it is good for the virtuous to be doing well and bad for them to be doing badly. And the fact that your benefactor is virtuous is manifest in your life: you are a little better off due to the fact that they helped you. So the value of their life is significant to you: you should be especially emotionally involved in how they are doing and have special reason to do them a little good. Reasons of gratitude, then, arise from contact with value. They arise when someone's virtue is manifest in your life.

Now we turn to compensation. Sometimes, regrettably, we wrong others. We fail to respect their claims. Plausibly, we should care more about our own wrongdoings than about the wrongdoings of other people. If you stole someone's car, you should be more troubled by this injustice than by similar thefts committed by strangers. And you have stronger reasons to compensate your victim than you have to compensate similar victims of others. Here is how the contact account can explain this. That your victim is unjustly lacking her car is a value fact: it makes it appropriate for us all to feel sorry for them and to give them their car if we can. Moreover, this fact manifests your dispositions: your lack of concern for property rights, say, is your disposition to wrongfully appropriate others' possessions, and the fact that your victim is unjustly lacking their car is a manifestation precisely of that. So you should care especially about this fact. You have extra reason to feel sorry for them and to hand them back their car. Reasons of reparation, then, arise from contact with value. They arise when your lack of moral concern is manifest in a wrongdoing.

These two cases raise a question. We say that according to the contact account, you have extra reason to help those you should compensate or show gratitude. But intuitively, you often do not merely have reason to help such

people: you have a *duty* to do so. You owe them gratitude or compensation. How do we get from reasons to this duty? Such a question also arises in the core cases of special relationships. Often, it is not just that we have special reasons to care for our friends or children: we have obligations to care for them. How do we get obligations out of such reasons? This question clearly is not one that only the contact account faces. It can be asked of anybody who takes a reasons-first approach to ethics.[28] Fortunately, the contact account seems compatible with any plausible answer to this question. One plausible answer, for example, derives obligations from the relative strength of reasons. It says that an action is morally obligatory if and only if the moral reasons that support it are stronger than all of the reasons—moral and nonmoral—that support any alternative.[29] On this view, contact with value gives you duties or obligations when it intensifies the strength of your moral reasons to do something to such an extent that they are stronger than your reasons to do otherwise.[30]

## 6. DISCOUNTING

Let us further strengthen the case for the contact account by applying it to some cases of nonmoral normativity—that is, different cases of discounting. A familiar example is *future discounting*: generally, we care less about things the more distant they are in the future. You are more excited about your trip to Mexico next week than about your trip to the Vatican in fifty years. You save more money now to spend on tomorrow's mole and Mayan ruins than you save to spend on far-off pizza and papal residencies. *Past discounting* is just as familiar: generally, we care less about things the more distant they are in the past. When a relative dies, when you are deeply wronged, or when one of your hopes is thwarted, you first feel intense grief, anger, or disappointment, and you may be motivated to do a lot about it. But then the intensity of your emotions wanes, and you are likely no longer as inclined to act.[31] This, it seems, is as it should be. It is bizarre to care as much about a trip in fifty years as about one that starts

28 For an especially thorough exploration of this view, see Schroeder, *Reasons First*. For more on this debate, see Portmore, *Commonsense Consequentialism*, ch. 5; Snedegar, "Reasons, Oughts, and Requirements"; and Schmidt, "How Reasons Determine Moral Requirements" and "The Balancing View of Ought."

29 This view is defended in de Kenessey, "The Relation Between Moral Reasons and Moral Requirement."

30 For more on this point, see Lovett and Riedener, "Commonsense Morality and Contact with Value," 426.

31 For prior discussion of this, see Marušić, "Do Reasons Expire" and *On the Temporality of Emotions*; Callard, "The Reason to Be Angry Forever"; and Na'aman, "The Rationality of Emotional Change."

next week. It is pathological to care unabated about past losses, injustices, or disappointments without regard for the passage of time. The contact account vindicates these phenomena too. It can let us understand these phenomena too as kinds of partiality.

Let us first see this with future discounting. Recall that on the contact account, what matters is whether your current life is manifest in a fact or whether a fact is manifest in your current life. Thus consider your near future. Many of your current dispositions will be manifest in the value of your upcoming weeks. The value of your Mexico trip will be grounded in the hikes and dives and parties you will engage in. And these will manifest your current adventurousness, impetuousness, or celebratory mood, so you should care a lot about them. Your far future may still manifest some of your current dispositions. The value of that distant Vatican trip might partly be grounded in your great future understanding of Renaissance art. And that understanding might be a distant manifestation of the curiosity that characterizes you already now. But you will change over time, lose many of your current dispositions, and acquire new ones instead. Generally, fewer of your current dispositions will be manifest in things the more distant they are in the future. So you should care about your near future more. But that is just to vindicate future discounting.

A parallel story applies to past discounting. Imagine your friend died yesterday, an untimely death. The badness of this death is manifest in your current life in many ways. It is manifest in all the absences that they leave: in the conversations gone silent with their demise, the chess games for which you now lack a partner. You have yet to fill the void in your life that their departure is disposed to cause. It is also manifest in your emotional response: you currently feel an extreme grief and sense of loss. And this itself is a manifestation of the terrible tragedy of this death. But both of these things will wane over time: you will get other friends, have other conversations and different chess partners. You will think of the friend's death less and less and will feel less intensely the pain of their loss. Generally, fewer dispositions related to a value fact will be manifest in your life the more distantly that fact lies in the past. So you should care about the near past more. And that is just to vindicate past discounting.

Finally, let us turn to a different, less familiar phenomenon: *modal discounting*. We treat close and remote possibilities differently. Imagine you are driving home from work. It is a wet and foggy night. Suddenly, a car comes from the other direction. They have gone around a corner too fast and are spinning out of control. You slam on the brakes and only narrowly avoid a crash. You could have died. Now of course, every time you get in your car, you could in principle die. But you care more about the modal fact when you narrowly avoid death than when death was only a distant option. More generally, we care more about

the fact that something could have happened when it almost did than when there was little chance of it happening. We care more about close than remote possibilities. Can the contact account vindicate such modal discounting?

We think that it can. The key point is that the grounds of close possibilities are manifest in your life to an extent that those of remote ones are not. Consider the fact that you could have died on your way home from work. When you almost died, many of the grounds of this fact are manifest in your life. The slipperiness of the road is manifest in your uncontrolled steering. The fogginess of the night is manifest in your hazy vision. The other car's causal powers are manifest in that guardrail near you being destroyed. When you were not at all close to dying, few grounds of the fact that you could have died are manifest in your life. Perhaps the facts that there were other drivers on the road or that you do not have lightning reflexes or an invulnerable body are manifest in your life. But these facts will be manifest in your life in the former case too. Thus, you are more in touch with the close possibility than the remote one. You should care about it more. And this just is to explain modal discounting.

In sum, the contact account has a very wide range of application. It explains the personal significance of people, objects, and projects. It explains the reasons that arise from gratitude and compensation. And it explains our reasons for temporal and modal discounting. All are assimilated to a kind of partiality. This is our master argument for the contact account of significance. Explanatorily speaking, it is enormously powerful. Any view that explains such a wide swathe of ethical phenomena is worth taking seriously indeed.

## 7. CONCLUSION

Let us conclude. We think the contact account of significance expresses a very intuitive idea. Contact with value matters. This idea is explanatorily powerful, and it was present in Western philosophy from its inception. In the *Republic*, Plato proposed that we should unshackle our chains, walk out of the Cave, and stare squarely at the shining light of goodness.[32] In the *Symposium*, he suggested that we should follow the sweet call of Eros, and beget good things.[33] The best life, he suggested, was the life in some kind of contact with the good. The contact account is reminiscent of Plato's view. But it focuses on personal significance rather than the good life. It says that contact with value intensifies the weight of our reasons to care about that value. Despite the difference, we take heart from the parallel between Plato's view and the contact account of

32 Plato, *The Republic*, 514a–520a.

33 Plato, *The Symposium*, 206a–212a.

significance. We think that both get at a magnetic but elusive idea. Both get at the sublimity of touching the good.[34]

*Australian Catholic University*
*adam.lovett@acu.edu.au*

*University of Bergen*
*stefan.riedener@uib.no*

## REFERENCES

Brandt, Josh. "Negative Partiality." *Journal of Moral Philosophy* 17, no. 1 (2020): 33–55.

Brighouse, Harry, and Adam Swift. "Legitimate Parental Partiality." *Philosophy and Public Affairs* 37, no. 1 (2009): 43–80.

Brink, David O. *Moral Realism and the Foundations of Ethics.* Cambridge University Press, 1989.

Callard, Agnes. "The Reason to Be Angry Forever." In *The Moral Psychology of Anger*, edited by Myisha Cherry and Owen Flanagan. Rowman and Littlefield, 2017.

Chang, Ruth. "Introduction." In *Incommensurability, Incomparability, and Practical Reason*, edited by Ruth Chang. Harvard University Press, 1997.

Choi, Shungho, and Michael Fara. "Dispositions." In *Stanford Encyclopedia of Philosophy* (Fall 2018). https://plato.stanford.edu/archives/fall2018/entries/dispositions/.

de Kenessey, Brendan. "The Relation Between Moral Reasons and Moral Requirement." *Erkenntnis* (2023): 1–22.

Drake, Nicholas. "Love, Reasons, and Desire." *Ethical Theory and Moral Practice* 23, no. 3 (2020): 591–605.

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

Frankfurt, Harry. *The Reasons of Love.* Princeton University Press, 2004.

Jeske, Diane. "Friendship, Virtue, and Impartiality." *Philosophy and Phenom-*

34 For helpful comments on our manuscript, we thank Alexander Dietz, Tyler John, Zoë A. Johnson King, Benjamin Lange, Jörg Löschke, J. McKenzie Alexander, Leo Menges, Jake Nebel, Peter Schaber, Samuel Scheffler, Philipp Schwind, Jake Zuehl, two anonymous reviewers for the *Journal of Ethics and Social Philosophy*, as well audiences in Bern and Zurich.

*enological Research* 57, no. 1 (1997): 51–72.
———. *Rationality and Moral Theory: How Intimacy Generates Reasons.* Routledge, 2008.
Jollimore, Troy A. *Love's Vision.* Princeton University Press, 2011.
Keas, Michael. "Systematizing the Theoretical Virtues." *Synthese* 195, no. 6 (2018): 2761–93.
Keller, Simon. *Partiality*. Princeton University Press, 2013.
Kirwin, Claire. "Value Realism and Idiosyncrasy." In *Oxford Studies in Metaethics*, vol. 18, edited by Russ Shafer-Landau. Oxford University Press, 2023.
Kolodny, Niko. "Which Relationships Justify Partiality? The Case of Parents and Children." *Philosophy and Public Affairs* 38, no. 1 (2010): 37–75.
———. "Which Relationships Justify Partiality? General Considerations and Problem Cases." In *Partiality and Impartiality: Morality, Special Relationships, and the Wider World*, edited by Brian Feltham and John Cottingham. Oxford University Press, 2010.
Kubala, Robbie. "Valuing and Believing Valuable." *Analysis* 77, no. 1 (2017): 59–65.
Lange, Benjamin, and Josh Brandt. "Partiality, Asymmetries, and Morality's Harmonious Propensity." *Philosophy and Phenomenological Research* 109, no. 1 (2023): 1–42.
Lazar, Seth. "The Justification of Associative Duties." *Journal of Moral Philosophy* 13, no. 1 (2016): 28–55.
Lewis, James H. P. "The Aesthetics of Coming to Know Someone." *Philosophical Studies* 180 (2023): 1675–690.
Lord, Errol. "Justifying Partiality." *Ethical Theory and Moral Practice* 19, no. 3 (2016): 569–90.
Löschke, Jörg. "Relationships as Indirect Intensifiers: Solving the Puzzle of Partiality." *European Journal of Philosophy* 26, no. 1 (2017): 390–410.
Lovett, Adam, and Stefan Riedener. "Commonsense Morality and Contact with Value." *Philosophy and Phenomenological Research* 1, no. 1 (2024): 1–21.
———. "The Good Life as the Life in Touch with the Good." *Philosophical Studies* 181, no. 5 (2024): 1141–65.
Marušić, Berislav. "Do Reasons Expire? An Essay on Grief." *Philosophers' Imprint* 18 (2018): 1–21.
———. *On the Temporality of Emotions: An Essay on Grief, Anger, and Love.* Oxford University Press, 2022.
Na'aman, Oded. "The Rationality of Emotional Change: Toward a Process View." *Noûs* 55, no. 2 (2021): 245–69.
Naar, Hichem. "Subject-Relative Reasons for Love." *Ratio* 30, no. 2 (2017): 197–214.

Olson, Nate W. "Review: Simon Keller, *Partiality*." *Ethics* 124, no. 3 (2014): 622–26.

Plato. *The Republic.* Edited by G. R. F. Ferrari. Translated by Tom Griffith. Cambridge University Press, 2000.

———. *The Symposium.* Edited by M. C. Howatson and Frisbee C. C. Sheffield. Translated by M. C. Howatson. Cambridge University Press, 2008.

Portmore, Douglas W. *Commonsense Consequentialism: Wherein Morality Meets Rationality*. Oxford University Press, 2011.

Raz, Joseph. "Liberating Duties." *Law and Philosophy* 8, no. 1 (1989): 3–21.

Rosen, Gideon. "Metaphysical Dependence: Grounding and Reduction." In *Modality: Metaphysics, Logic and Epistemology*, edited by Bob Hale and Aviv Hoffmann. Oxford University Press, 2010.

Scheffler, Samuel. *Boundaries and Allegiances: Problems of Justice and Responsibility in Liberal Thought.* Oxford University Press, 2002.

———. "Membership and Political Obligation." *Journal of Political Philosophy* 26, no. 1 (2018): 3–23.

———. "Morality and Reasonable Partiality." In *Partiality and Impartiality: Morality, Special Relationships, and the Wider World*, edited by Brian Feltham and John Cottingham. Oxford University Press, 2010.

———. "Projects, Relationships, and Reasons." In *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, edited by R. Jay Wallace. Oxford University Press, 2004.

———. "Relationships and Responsibilities." *Philosophy and Public Affairs* 26, no. 3 (1997): 189–209.

———. "Valuing." In *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, edited by R. Jay Wallace, Rahul Kumar, and Samuel Freeman. Oxford University Press, 2011.

Schmidt, Thomas. "The Balancing View of Ought." *Ethics* 13, no. 2 (2023): 246–67.

———. "How Reasons Determine Moral Requirements." In *Oxford Studies in Metaethics*, vol. 18, edited by Russ Shafer-Landau. Oxford University Press, 2023.

Schroeder, Mark. *Reasons First.* Oxford University Press, 2021.

Seglow, Jonathan. *Defending Associative Duties*. Routledge, 2013.

Seidman, Jeffrey. "Valuing and Caring." *Theoria* 75, no. 4 (2009): 272–303.

Setiya, Kieran. "Other People." In *Rethinking the Value of Humanity*, edited by Sarah Buss and Nandi Theunissen. Oxford University Press, 2023.

Snedegar, Justin. "Reasons, Oughts, and Requirements." In *Oxford Studies in Metaethics*, vol. 11, edited by Russ Shafer-Landau. Oxford University Press, 2016.

Stroud, Sarah. "Permissible Partiality, Projects, and Plural Agency." In *Partiality and Impartiality: Morality, Special Relationships, and the Wider World*, edited by Brian Feltham and John Cottingham. Oxford University Press, 2010.

Velleman, J. David. "Love as a Moral Emotion." *Ethics* 109, no. 2 (1999): 338–74.

Williams, Bernard. "Internal and External Reasons." In *Moral Luck*. Cambridge University Press, 1981.

———. "Persons, Character, and Morality." In *Moral Luck*. Cambridge University Press, 1981.

*Journal of Ethics and Social Philosophy*
VOL. 30, NO. 2 · APRIL 2025
*https://doi.org/10.26556/jesp.v30i2.4064*
© 2025 Author

# A FIDUCIARY THEORY OF PROPERTY

*Rutger Claassen*

PROPERTY RIGHTS do not just give a person the power to rule over things, for example to command the use of the material or immaterial assets that the person holds as their own. They also give an owner power over other persons. An owner's legally granted authority to determine who can enter their home, for example, implies a power to exclude others from enjoyment of the same good. Property is always a triadic relation, between an owner, a thing, and a nonowner. This is more or less a truism in property theory. But some owners have control over vital resources that are of great importance to the lives of others. Hence ownership power can become harmful to nonowners, more or less analogous to the way the exercise of political power can become problematic for those subjected to it. This analogy between property and political power received canonical expression in the work of early twentieth century American legal realists such as Morris Cohen, who claimed that "dominium over things is also imperium over our fellow human beings."[1] However, the analogy has always remained suggestive at best, and it has never been accepted by standard libertarian or egalitarian approaches to property. Can it be saved?

The strategy that I explore in this paper is to apply to ownership a widespread understanding of political power in the Western tradition: as an *office* in which governments hold their power in trust, to the benefit of those over whom they rule.[2] This is a *fiduciary* understanding of the political relation between rulers and subjects. In such a fiduciary relation, a fiduciary is empowered to make decisions over a beneficiary, for the interest of that beneficiary. In the private sphere, lawyer-client, doctor-patient, and teacher-student relations are prominent examples of fiduciary relations. The present paper proposes to apply this fiduciary model to ownership as well. It argues that owners have fiduciary duties to nonowners when they interact with them on terms that risk exploiting them. When properly enforced, these fiduciary duties tame the power of

1 Cohen, "Property and Sovereignty," 13. More recently, the emphasis on power has been retrieved in property theory by Robé, *Property, Power and Politics*; and Ireland, *Property in Contemporary Capitalism*. For a response to Robé, see Claassen and Katz, "Property."

2 Fox-Decent, *Sovereignty's Promise* and "Challenges to Public Fiduciary Theory"; Leib et al., "Mapping Public Fiduciary Relationships"; and Criddle et al., *Fiduciary Government*.

ownership and prevent abuses of power over nonowners. With great power comes great responsibility—in this, political rulers and owners should be alike.

The paper is set up as follows. In section 1, I argue that libertarian and egalitarian property theories, while opposite in many ways, share an *absolutist* understanding of ownership. *Stewardship* theories of property reject this understanding and position the owner as responsible to others. The fiduciary property theory is a species of such a stewardship view.

In sections 2 to 5, I develop my fiduciary property theory. In section 2, I build on Larissa Katz's theory of ownership as an office and argue that ownership is a legally created power that makes owners into fiduciaries for the legal system as a whole. In section 3, I argue that owners become fiduciaries for particular nonowning persons, in two types of situations: when they cooperate with others in contexts of incomplete contracts, and when they impose externalities on others. In both contexts, the risk of exploitation justifies the imposition of fiduciary duties. In section 4, I elaborate a conception of exploitation to specify and justify this position. Exploitation happens when nonowners are unfairly taken advantage of; it requires the acceptance of a specific theory of justice. By avoiding exploitation, owners contribute to realizing a just society. In section 5, I turn to the problem of how to implement fiduciary duties in practice. Both court enforcement and democratization of ownership are ways of dealing with this problem.

The final two sections compare my fiduciary property theory to relevant alternatives. In section 6, I contrast it with the work of three property theorists who have proposed similar views. In section 7, I return to the contrast between fiduciary and egalitarian property theories. Both promise to restrain the powers of owners over nonowners, but in different ways. The fiduciary relationship holds the potential for reforming the relationship between public authorities and private owners, from an antagonistic to a (more) cooperative one.

## 1. LOCATING FIDUCIARY VIEWS WITHIN PROPERTY THEORY

In this section, I sketch the landscape of property theory with a view to locating how a fiduciary theory fits into it. Simplifying a rich tradition that cannot be comprehensively surveyed here, I start with the basic dichotomy between libertarian and (liberal-)egalitarian views on property.[3] Libertarian property theories defend robust property rights. They often (though not always) are philosophical descendants of Locke's property theory, which allows for

3 For surveys of the tradition, see, e.g., Garnsey, *Thinking About Property*; Pierson, *Just Property*, vols. 1–3; Becker, *Property Rights*; and Ryan, *Property and Political Theory*.

unilateral appropriation under the right conditions.[4] Property rights on a libertarian view are meant as a protection against the state encroaching on an owner's power over their goods. Hence the bar for state regulation must be appropriately high. Tomasi expresses this in a Rawlsian framework by arguing that all property rights should be counted among the basic liberties, which can be restricted only for the sake of other basic liberties.[5]

Egalitarian property theories locate themselves opposite to libertarian property theory. They see property rights as state-guaranteed rights that are *instrumental* to a higher-order (liberal) good. For example, Rawls designs his two principles of justice so as to realize for each citizen the two moral powers, their capacity for a sense of justice and their capacity for a conception of the good. In light of this normative justification, property rights in personal goods are classified as basic liberties, but property rights in the means of production are not.[6] Hegelian and Kantian theories are another example. Hegelians see the value of ownership in the opportunities for owners to express their will in the control over their properties.[7] For Kantians, property can be justified only as an expression of the omnilateral will, in contrast to Lockean unilateral appropriation.[8] Such theories can be developed in an egalitarian direction, arguing for regulation or redistribution of property rights.[9]

To locate fiduciary views of property, we must first understand what libertarian and egalitarian views have in common, despite all their differences. Both accept that ownership entails having discretionary power over goods, within the limits established by law. Beyond these legal limits, owners rule absolutely in that they are *absolved* from justifying their choices to third parties, as a public official normally is required to do.[10] As Hanoch Dagan says, "When public officials occupy a position of authority, they purport to represent the state. . . . The authority of owners, by contrast, relies on their subjectivity—their intention,

4 Locke, *Two Treatises of Government*; Nozick, *Anarchy, State, and Utopia*; and Narveson, *The Libertarian Idea*.

5 Tomasi, *Free Market Fairness*. Similarly, see Gaus and Lomasky, "Are Property Rights Problematic?"

6 Rawls, *A Theory of Justice*.

7 Hegel, *Elements of the Philosophy of Right*.

8 Kant, "The Metaphysics of Morals."

9 See Waldron, *The Right to Private Property*; Radin, "Property and Personhood"; and Ripstein, *Force and Freedom*.

10 Katz, "Property's Sovereignty," 304. The notion of absolutist property is often associated with Blackstone's famous dictum of ownership as a "sole and despotic dominium." See Blackstone, *Commentaries on the Laws of England*, bk. 2, p. 1.

judgment, and point of view—as a free-standing source of legal claims over others."[11]

This absolutist nature of property does not mean that an owner's power is unlimited. The qualifier 'within the limits of the law' is crucial. Public laws can restrict specific exercises of property rights by imposing duties in the public interest. For example, a law can restrict my right to trade my properties by specifying times and places where trade can take place. Throughout this paper, I will refer to such state-imposed legal duties as *regulatory duties*, to avoid confusion with fiduciary duties (which can also be imposed by law). Absoluteness pertains to the decision space left to owners *after* regulatory duties have been imposed. Libertarian positions (pleading in favor of minimizing legal restrictions on ownership) as well as egalitarian positions (pleading in favor of more substantive restrictions) differ on the size of these legal restrictions but agree on the underlying absolutist conception of property. This makes the absolutist conception virtually unchallenged in modern property theory.[12]

Fiduciary property theories reject this absolutist conception of property. A fiduciary relation typically involves two persons, a beneficiary and a fiduciary. Between them, there is a fiduciary relation when: (1) the fiduciary exercises discretionary power over the interests of a beneficiary; (2) the beneficiary is vulnerable to (or dependent on) the fiduciary; and (3) the beneficiary must trust the fiduciary to act in their best interests, and the fiduciary must act in a trustworthy manner.[13]

11 Dagan, *A Liberal Theory of Property*, 62.

12 Olsen, "The Early Modern 'Creation' of Property and Its Enduring Influence," 127; and di Robilant, *The Making of Modern Property*, 29. Absolutist notions of ownership are often associated with classical liberalism, against which egalitarians would hold a "bundle of rights" view, on which several persons can hold various "incidents of ownership" with respect to the same good. See Honoré, "Ownership." But on my understanding, absolutism does not necessarily imply "full liberal ownership." For example, a landlord holds rights to sell their property and retains the right to get an income but has contracted away the right to use the property to a tenant. These fragmented property constructions are simply a matter of legal reality. See Grey, "The Disintegration of Property." The point is that *both landlords and tenants can be absolutist* with respect to the part of the bundle of property rights that each of them holds. No fiduciary orientation necessarily comes into being just by dividing the incidents of the bundle over multiple persons. I will for simplicity speak of owners but mean to include holders of lesser property rights as well.

13 Frankel, *Fiduciary Law*, 2; Miller, "The Fiduciary Relationship," 69; and Fox-Decent, *Sovereignty's Promise*, 29. There are competing legal theories about how to define and understand the relation. For an introduction, see Smith, "Parenthood Is a Fiduciary Relationship," 402–18.

A fiduciary duty is, in the first instance, a moral duty. There is a separate field of "fiduciary ethics."[14] But as is the case for other types of moral duties, a subclass of them can be recognized and enforced in law. Under fiduciary law, fiduciaries have two main types of duties. The duty of loyalty requires the fiduciary to act as a loyal representative of the beneficiary's interests.[15] This duty would be breached if the fiduciary abuses the beneficiary's trust for their own interest. Conflicts between the fiduciary's own interest and the interests of any beneficiaries must be avoided.[16] The duty of care refers to a reasonable standard of care that the fiduciary must live up to. If the fiduciary is not sufficiently careful, she acts with negligence.[17]

According to many authors, the duty of loyalty is at the heart of the fiduciary relationship. It expresses a demand that can be cashed out in subtly different ways. Generalizing over these differences, the crucial point is that loyalty does not require one particular course of action but instead refers to a particular attitude or orientation on the part of the fiduciary. Since the power is discretionary, no particular course of action can be established as "the right one," independent from the judgment of the fiduciary. The fiduciary, in the terminology of Stephen Galoob and Ethan Leib, must have an intention to "attribute nonderivative significance to her beneficiary's interests."[18] Or it requires, in Lionel Smith's words, that the fiduciary "exercise judgment in what they subjectively believe to be the best interests of the beneficiary."[19] Or finally, in Daniel Viehoff's terms, it requires the fiduciary to deliberate in such a way that she is "guided by" the interests of the beneficiary.[20] In brief, the fiduciary must engage in an internal decision-making process resulting in an intention (judgment/decision) to act on behalf of someone else; and this process has as its focal point the interests of that other person.

What would it mean to understand ownership as a fiduciary relation? Interestingly, property has been at the heart of the development of fiduciary law, so

14 Mussell, "Theorising the Fiduciary."

15 Elsewhere I discuss the representative nature of the property relation, in the context of work on representation in democratic theory. See Claassen, "Property as Power."

16 Samet, *Equity*, 114.

17 Smith, "Parenthood Is a Fiduciairy Relationship," 200–1. What both duties imply in practice may depend. The duties of parents, teachers, lawyers, doctors, etc. are each determined by the very different relationships they have with their beneficiaries. Moreover, the duties depend upon the wider culture's ideas about good teaching, doctoring, raising children, etc. For ownership, a separate account is necessary, as for any other fiduciary relation.

18 Galoob and Leib, "Intentions, Compliance, and Fiduciary Obligations," 10.

19 Smith, "Fiduciary Relationships," 611.

20 Viehoff, "Legitimate Injustice and Acting for Others," 332.

much so that fiduciary theorists feel a need to argue that "fiduciary relationships are not confined to situations of financial or property management."[21] However, property and fiduciary relations meet each other here to regulate situations in which a fiduciary manages someone else's property, e.g., when she acts as a trustee, the director of a company, or a legal advisor. Usually, the beneficiary has voluntarily trusted their property to the fiduciary for this purpose. The vulnerability of the beneficiary is self-inflicted. The fiduciary property theory to be developed in this paper, however, is concerned with a very different phenomenon: the situation where the fiduciary manages *their own* property, and the beneficiary is a *nonowner* who has not voluntarily consented to anything but experiences the negative consequences of the owner's exercise of ownership rights. She is a *negative beneficiary*, so to speak. Fiduciary duties in such situations have as their purpose to protect these vulnerable nonowners.[22]

This application of fiduciary principles to property, which is not recognized in fiduciary law, is what I will refer to as *fiduciary property theory*. It requires a series of further explanations, which this paper aims to provide. But first: Are there any historical examples of fiduciary property theories? To the best of my knowledge, fiduciary property theories have had a marginal presence in property theory overall. They are known under the label of "stewardship" theories and have mainly been defended in two contexts.

First, various religious and Indigenous traditions hold stewardship views of property.[23] For example, Catholic social doctrine declares that the main principle for thinking about property is the *universal destination of goods*: owners are to use their property so as to also benefit the common good.[24] In the Christian tradition, this view goes back to Aquinas and Aristotle, who argued that private

21 Smith, "Parenthood Is a Fiduciairy Relationship," 452.

22 Note that it is certainly not uncommon that fiduciary relations are recognized where there is no consent to transfer authority from the (vulnerable) beneficiary to the fiduciary: the incapacity of a child or a comatose patient are examples of such situations. Nonowners here are in a similar position.

23 For example, Dagan writes about the Jewish conception of property: "Property was entrusted to us by God for our well-being; but property-owners are mere custodians of the resources they hold" (*Unjust Enrichtment*, 58). The Islamic view of property starts from a similar position. See Sait and Lim, *Land, Law, and Islam*. For a comparative discussion of ownership duties in Jewish, Christian, and Islam traditions, see Singer, *The Edges of the Field*. For stewardship in Indigenous traditions, see Carpenter et al., "In Defense of Property."

24 For the latest version of the Catholic social doctrine, see Pontifical Council for Justice and Peace, *Compendium of the Social Doctrine of the Church*, sec. 178.

ownership is always burdened with a duty to "common use."[25] Practically, the most important application is that owners have duties to share their superfluous possessions with those in need. This grounds the Christian duty to practice charity.[26]

A second context in which stewardship theories play a role is environmental ethics and environmental law. Here, some authors argue that owners should act as stewards for future generations and possibly also for animals and ecosystems.[27] This would justify concrete duties to act as responsible owners with respect to land and natural resources, in particular with the aim of conservation. Owners here are put in a double position: they are both themselves beneficiaries of these resources and charged with the task of taking into account the legitimate interests of other beneficiaries, like future generations.

I do not have the space for a full discussion doing justice to these theories. One obvious problem for religious property doctrines is that they rely on spiritual foundations, which are not easily acceptable to all in a secular context. The environmental notion of stewardship is restricted in the scope of its application to land and natural resources. Can a fiduciary property theory be constructed with a broader scope (applicable to all types of resources) *and* with a justification that is acceptable in the context of a liberal-democratic society? Or is a fiduciary approach an antiquated dead end in property theory, as many seem to believe?[28]

25 Aquinas builds on canon lawyers' treatments of the duty to share with the poor in situations of dire need. See Tierney, *The Idea of Natural Rights*, 1150–625, 70–76. This "right of necessity" was shared by later thinkers such as Grotius, Pufendorf, and Locke. Aristotle, like the later Christian tradition, struck a middle road between unconstrained private property and communal property, in which private owners must accept a condition of holding property "common in use." See Aristotle, "Politics." For discussion, see Lametti, "The Objects of Virtue," 5–15; and Frank, *A Democracy of Distinction*, 54–80. For a contemporary view building on both Aristotle and Aquinas, see Lametti, "The Concept of Property."

26 Aquinas, *Political Writings*, 208; and Finnis, *Aquinas*, 188–96.

27 See Attfield, *The Ethics of the Global Environment*; Lees, "Property in the Anthropocene"; and Rodgers, "Nature's Place?"

28 I say antiquated since some critics associate fiduciary property views with feudalism. For example, legal scholars criticizing Katz's and Essert's theories of ownership as office (on which I rely hereafter) have characterized the idea of office as amounting to bringing "premodern" or "feudal" property forms back in. See Ripstein, "Property and Sovereignty," 254; and Penner, "Private Law Offices," 312. For Katz's reply to this, see Katz, "It's Not Personal," 112. In my view, the association of a fiduciary property theory with feudalism is not wrong—feudalism did conceive of property as trusteeship. See Rose, "Property as Wealth, Property as Propriety," 237. The question is whether it is necessarily feudal or whether a nonfeudal, liberal-egalitarian fiduciary theory can be developed. I discuss the relation between modern property and feudalism in Claassen, "Property and Political Power."

## 2. OWNERSHIP AS AN OFFICE

Fiduciary relations lead to the creation of an office, which the fiduciary holds on behalf of others.[29] Recently, two legal scholars have proposed a theory of ownership as an office. I here focus on Larissa Katz's elaboration of this idea.[30]

Katz first introduces an uncontroversial understanding of offices. They are positions of authority, roles to be filled by office holders in the context of a wider practice, such as a religious practice (a priest), a familial practice (a parent), or a political practice (a president). The role is separable from the holder—i.e., it can be fulfilled by different people acceding to the office in turn. This makes the exercise of authority to some extent impersonal. Finally, the position is a stable, enduring source of authority because when an officer dies or resigns, the office can be filled by someone else. Orphans need new, legally appointed parents, as presidencies need presidents, and companies need directors.[31] Offices are an institutionally embedded form of power, transcending the individuals fulfilling the role.

Katz applies this notion to ownership. For her, the core of ownership is "the special authority to set the agenda for a resource."[32] This definition hints at an analogy with sovereignty, the exclusive agenda-setting authority of those occupying the office of government.[33] The authority of owners is the result of a form of delegation of authority by the state. We can think of this as a constitutional choice, on a par with delegations of authority to towns and provinces or other lower levels of government. Owners are licensed to do things with their property that nonowners are precluded from doing. According to Katz, a primary argument for conceiving of ownership as an office is that it captures the *impersonality* of ownership. Office holders' authority must be recognized by all others, in virtue of their holding of the office, regardless of the specific agreements (contracts) between the office holder and these others. Ownership rights are *in rem*, not *in personam*, and the notion of office explains what that means. The authority towards third parties is linked to the owner's authority over things. When ownership of a thing is transferred to someone else, then

29 For the close link between fiduciary relations and the notion of office, see Dagan and Scott, "Reinterpreting the Status-Contract Divide"; and Samet, *Equity*, 124.

30 Katz has written extensively on the topic. See Katz, "Property's Sovereignty," "It's Not Personal," and "Ownership and Offices." Hence, I focus on her work. Essert wrote one article defending the idea ("The Office of Ownership") but later expressed skepticism ("The Office of Ownership Revisited").

31 Katz, "Governing Through Owners," 2037.

32 Katz, "Property's Sovereignty," 316.

33 Katz, "Exclusion and Exclusivity in Property Law," 293–95.

all others acquire a duty to accept the authority of the new owner. Nonowners have duties to respect the authority of the owner, regardless of the specific identity of the owning person at any point in time.[34]

Katz's understanding of office is formal. It accepts the absolutist action orientation of owners as providing the baseline for their decisions. The unique feature of Katz's ownership-as-office theory is to interpret this absolutist power as a matter of holding office. This combination of absolutist ownership with the idea of office has been criticized by Arthur Ripstein, who argues that an office must be understood more substantively, as a position in which you act as a steward on behalf of others: "Ordinarily, an official is charged with advancing or protecting the purposes of the institution in which that office is found. By contrast, an owner typically has untrammeled discretion with respect to the purposes for which the property will be used. 'Do whatever you want' is not a mandate."[35] It is characteristic of offices, on a substantive understanding, that they must be fulfilled to best realize the interests of the persons over whom authority is exercised. This is office as the locus for a *fiduciary* relation.

So should we go beyond Katz's formal concept of office and conceive of ownership as such a more substantive, fiduciary office? Two initial objections may seem to make this a problematic move.

The first is that fiduciary relations normally apply between one fiduciary and one concrete beneficiary party, on the model of binary relations such as doctor-patient and teacher-student. Is the generic power of ownership, holding *erga omnes*, too diffuse to be qualified as fiduciary? Notably, some other offices have as their task not to represent a designated set of particular beneficiaries but a purpose. Miller and Gold refer to these as situations of *fiduciary governance*.[36] For example, a judge holds an office but does not represent the concrete interests of (one of) the two litigating parties in front of them; that is the lawyers' job. But judges do make authoritative decisions, and their authority is impersonal and transmissible to other judges. They fulfill a purpose worth fulfilling in the eye of the state that creates the office.[37] The recent application of fiduciary theo-

34 In addition, Katz claims the notion of office also helps to understand the puzzle of the *transmissibility* of ownership. A transfer of ownership can be understood as the succession of a new officeholder to the office. See Essert, "The Office of Ownership," 436; and Katz, "Ownership and Offices," 269. Finally, it helps us to understand why the law abhors *vacancies* in ownership. Unclarity about "who is in charge" upsets the legal order, for property management as much as for the governance of political territories. See Katz, "Ownership and Offices," 279.

35 Ripstein, "Property and Sovereignty," 254. Similarly, see Penner, "Private Law Offices," 313.

36 Miller and Gold, "Fiduciary Governance."

37 Katz, "Ownership and Offices," 274.

ries to political and public offices would be unthinkable without this extension of fiduciary duties from a binary duty to a duty discharged on behalf of multiple beneficiaries ("the public"), as defined by a purpose.[38]

The decision-making power of owners can also be understood as a matter of fiduciary governance, although an owner's office is obviously very different from that of public officials like judges. Indeed, Katz's formal characterization of ownership as an office to set the agenda for the use of things includes such a purpose. The purpose, or public interest, at stake, which, in her view, justifies the state's authorization of owners, is that of *coordination*: of creating a well-ordered society, minimizing the risk of conflicts, chaos, and violence that might take place when it is not specified who has an ownership claim over what.[39] Owners have a constitutionally mandated duty to uphold this scheme. This can be understood as an office they hold in trust, as fiduciaries of the legal system (and through it, of society at large).[40] To this extent, Katz's theory is fiduciary in a minimalist sense.

A second objection might be motivated by the widespread belief that the duty of loyalty forbids self-serving behavior. Judges are not authorized to decide cases in ways that serve their own personal interests. The same seems true for other private and public offices. But owners make self-serving decisions all the time. Does this not disqualify them from being fiduciaries of society? However, the prohibition on self-servingness must be qualified; and this is especially important in the context of fiduciary governance for a purpose, towards a wider group of people. For the fiduciary can be among the beneficiaries as well. A member of parliament must represent a multiplicity of constituents. The duty of loyalty is directed to all of them at once and requires the member of parliament to compromise between them on a fair and reasonable basis. And the member of parliament is herself one of these persons, for she is herself a resident of the district she represents.[41] Similarly, if five business partners appoint one of themselves as managing partner, that partner makes

38 Fox-Decent, *Sovereignty's Promise* and "Challenges to Public Fiduciary Theory"; Leib et al., "Mapping Public Fiduciary Relationships"; and Criddle et al., *Fiduciary Government*.

39 Katz, "Spite and Extortion," 1478. Elsewhere, Katz specifies the purpose in terms of a division of labor (in her discussion of Duguit's property theory): "The license to make self-regarding decisions was simply a matter of a convenient division of labour: there is social value in everyone's flourishing and it just so happens that the burden of making sure that I too flourish is delegated from society to me" ("It's Not Personal," 107). Essert mentions a range of normative purposes but remains agnostic ("The Office of Ownership," 437).

40 A similar suggestion has been made in republican political theory: Domènech and Bertomeu, "Property, Freedom and Money"; and Cassasas and Mundó, "Property as a Fiduciary Relationship and the Extension of Economic Democracy."

41 Criddle, "Stakeholder Fiduciaries," 125.

decisions on behalf of all of them, *herself included*.[42] In such contexts, the fiduciary needs to balance her own interests against those of others. This could be applied to ownership as well.

In conclusion, the notion of ownership as a formal office sets up the owner in a socially responsible role with a general fiduciary duty to do their part in the coordinating the use of goods in society. But this formal conception leaves the agenda owners set for their things completely up to the owners themselves; they may fully focus on self-regarding aims. The formal conception does not introduce particular fiduciary duties towards concrete nonowners as an actual restraint on the owners' absolutist discretion. It does not require introducing such additional duties, but it is compatible with such a move.[43] Then we would move to a substantive concept of office. When—if at all—would that be justified?

## 3. Triggering Fiduciary Duties: Incompleteness in and Beyond Contract

A first step in answering this question is to see when and why fiduciary relations normally emerge. Daniel Markovits offered an enlightening explanation in his discussion of the difference between contractual and fiduciary duties. In this section, I will show that ownership meets his description of when fiduciary duties arise.

Markovits's starting point is that contracts are the basic tool to facilitate cooperation between private parties. Legally, contractors are held to a duty to deal with each other *in good faith*, e.g., by not manipulating and deceiving one's contract partner. This sets a lower bar and allows for self-seekingness in all other respects. When a conflict arises after the contract has been made, then the intentions as laid down in the contract are the "lodestar" in resolving the problem. Markovits calls this a form of "sharing *ex ante*": the contract lays out how to share the burdens of a problem that emerges later or how to resolve a dispute about the distribution of benefits that emerge later.[44]

The contractual relation can be contrasted with the fiduciary relation. As Markovits explains, "a fiduciary relation becomes appealing partly because a principal requires her agent to act in ways that she cannot substantially specify

42 Criddle, "Stakeholder Fiduciaries," 110–17. Similarly, see Viehoff's discussion of "acting for others," in which he emphasizes that officials who act for citizens as beneficiaries are themselves among the beneficiaries: Viehoff, "Legitimate Injustice and Acting for Others," 358, 363.

43 Katz acknowledges this. See my discussion in section 6 below.

44 Markovits, "Sharing *Ex Ante* and Sharing *Ex Post*," 213.

*ex ante* or cannot directly evaluate *ex post.*"[45] These are situations in which the contract is necessarily incomplete (because of the impossibility to specify everything), and this incompleteness leaves the principal (beneficiary) vulnerable. There is a potentially abusive power relation.[46] As a solution, the contract partners decide to let the sharing decisions be made *ex post*: the fiduciary. The beneficiary "buys her fiduciary's initiative."[47] To safeguard the beneficiary's interests, the law imposes duties of care and loyalty on the fiduciary.

We can extend Markovits's insight. Any scheme of cooperation can suffer from the problem of contractual incompleteness. In the case of a society, John Rawls argues that cooperation raises a question of justice—the fair division of the benefits and burdens of social cooperation.[48] This will lead to the adoption of some highly abstract guidelines *ex ante* (the principles of justice); but the actual work to implement them is left to a government. In the case of a corporation, Luigi Zingales argues that cooperation between various stakeholders creates a need for corporate governance, in which decisions about the distribution of the rents are made *ex post*. Again, the need for governance is dictated by the fact that the original contract cannot specify in advance all the possible contingencies that may arise during the cooperation.[49]

Hence for both societies and corporations, the creation of a position of authority deciding on an ongoing basis about matters pertaining to the scheme of cooperation is justified by the impossibility of foreseeing everything in the initial contract. States and corporations manage property for their purposes. But the same thought can be applied to individual owners as well. For example, when involved in a series of ongoing transactions with a supplier, an owner may create expectations about future contracts that, when violated, are costly given the investments the supplier has already made. Similarly, temporary workers may invest in their skills because of the nature of the job they hold (so-called *asset-specific skills*) and be harmed when the employer does not extend the contract. Such an investment creates a risk of being vulnerable to the owner's decision once the investment has been made.[50]

Incompleteness of contract is so pervasive that we may wonder when the recognition of a fiduciary duty would be justified. What is the norm that triggers the shift in register from contractual to fiduciary duties? Markovits hints

45 Markovits, "Sharing *Ex Ante* and Sharing *Ex Post*," 215.

46 Markovits, "Sharing *Ex Ante* and Sharing *Ex Post*," 215.

47 Markovits, "Sharing *Ex Ante* and Sharing *Ex Post*," 216.

48 Rawls, *A Theory of Justice*.

49 Zingales, "Corporate Governance," 2325.

50 Scott, "The Law and Economics of Incomplete Contracts."

at the relevant norm when he argues that a breach of fiduciary duties happens when there is an abuse of power. I will refer to this phenomenon as *exploitation*, broadly understood as one party unfairly taking advantage of another. In the next section, I will elaborate a suitable notion of exploitation. For now, the core idea is simple: relations in which there is a sufficiently serious risk of exploitation trigger and justify fiduciary duties.

Beside incompleteness of contract, there is a second context in which fiduciary duties to specific parties may arise for owners: when they are not engaged in a voluntary scheme of cooperation but still interact with others in the sense that they impose an involuntary effect upon another person. In economics, these situations are referred to as *externalities*.[51] The link with property is made explicit by Demsetz in his economic theory of property rights. He argues that property rights allow societies to internalize externalities.[52] In the absence of property rights, one person's action very often imposes an externality on others. In connection to resources, one person using a resource diminishes what is available for others. In a state of abundance, this is no reason to worry. Property rights emerge where externalities become significant, and the benefits of having property rights outweigh the transaction costs involved with such a system.

However, the introduction of a property rights system often does not internalize all externalities. A persisting externality points to an incompleteness in the property rights system. Such a situation leads to a problem parallel to that of incomplete contracts. Why do externalities persist? Property rights fulfill the function of boundary setting, demarcating where the scope of one person's property right ends and the rights of others begin. The norms of property law help establish the boundaries between owners' property and nonowners' rights, determining what counts as a relevant externality and what does not.[53] But these norms are often not sufficiently detailed. Hence, disputes may arise, and there is no predetermined notion of property that will settle them.[54]

Here, too, incompleteness provides an occasion for the exercise of exploitative power by owners over nonowners. Here, too, a fiduciary duty is needed to fill the gap, tame this power, and prevent exploitation. To mention only the most obvious example, environmental pollution is often conceptualized as an

51 For in-depth discussion of the concept, see Hausman, "When Jack and Jill Make a Deal."

52 Demsetz, "Toward a Theory of Property Rights," 350.

53 Singer, "How Property Norms Construct the Externalities of Ownership," 60.

54 As Coase argues in "The Problem of Social Cost," externalities are about symmetric situations. When you have the right to build a skyscraper on your land, and that skyscraper blocks my sunlight, you impose an externality on me. When your landownership does not entail this right, then my right to prohibit you from building the skyscraper imposes an externality on you.

externality. The exploited parties (such as animals or future generations who are dependent on the environment) suffer from a gap in the property rights system. A fiduciary theory of property understands this as a situation in which a fiduciary duty should be imposed on owners. The legitimacy of ownership rights in environmental resources cannot be disentangled from the simultaneous presence of fiduciary duties to prevent the exploitative use of these rights with respect to these third parties.

The analysis in this section hence points to situations, in and beyond contract, in which owners can leverage their power over vulnerable nonowners and exploit them. These situations trigger a fiduciary duty not to exploit. But to make this analysis work, we need a conception of exploitation that underlies and justifies such fiduciary duties. Proposing such a conception is the task of the next section.

## 4. THE DUTY NOT TO EXPLOIT

The general concept of exploitation captures the wrong of situations where person *A* benefits from taking advantage of person *B*. The concept can be applied to both consensual and nonconsensual interactions, although it is often thought most pertinent to diagnose consensual interactions (since nonconsensual interactions are already considered wrong in virtue of the coercion involved). This broad range of application is important for ownership exploitation, which includes both types of situations (with externalities being nonconsensual but not obviously coercive).[55] The general concept of exploitation is broad in another sense as well: it can be applied to both mutually beneficial interactions and interactions that are harmful for the exploited party, although, for similar reasons, mutually beneficial interactions are often taken to be most interesting. In the following, I will work with examples of consensual and mutually beneficial interactions.

The crucial task in elaborating a theory of exploitation is to specify the condition of *taking advantage* and justify its wrongfulness. The condition is often specified as taking *unfair* advantage, since many authors argue we need a theory of justice to spell out which distribution of benefits is justified.[56] After all, a winning sports team benefits by taking advantage of the weaknesses of

55 Remember that externalities are symmetrical. (See previous note.) Hence, prior to a normative determination regarding which side should bear the externality, we cannot say who is coercing who. And once we have determined this, the coercion is justified (or the term 'coercion' does not apply).

56 There are, however, various theories of exploitation that explicitly reject such a basis in a theory of justice. For an example, see Vrousalis, "Exploitation, Vulnerability, and Social

their losing opponent, and these gains are legitimate, unless for some specific reason the losing team does not get what they are entitled to. Specifying these entitlements is the task of a theory of justice.

One prominent theory of exploitation uses the perfectly competitive market as a benchmark. Transactions in the context of a perfectly functioning market are fair (because reflecting the scarcity value of the good in question) and hence nonexploitative. By contrast, it is exploitative for someone to take advantage of another by exploiting a market imperfection, such as an information asymmetry.[57] On this view, a water-deprived, dehydrated person in the desert who arrives at the doorstep of the owner of the only water well is exploited by the latter if the owner sets an above-market price for their water. The owner exploits the fact that they are a (local) monopolist.

The main reason for rejecting such a market-based theory is that exploitation is not restricted to imperfect markets.[58] Imagine a desert example in which we do not have a monopolist but a set of competing owners of water wells. If they all charge a competitive market price of twenty dollars to the thirsty person, who is also a poor person and cannot afford that price, they are still exploiting that person. Admittedly, when drawing this conclusion, we rely on an egalitarian conception of justice, which moves beyond the identification of justice with whatever perfect markets happen to deliver. For the sake of the argument here, think of such an egalitarian theory of justice as holding that everyone is entitled to a basket of basic resources or to the fulfilment of basic needs or basic/central capabilities.[59] The thought then is that producers in perfect markets who leave consumers without access to these basic resources, needs, or capabilities act exploitatively.

In this paper, I presuppose the correctness of such an egalitarian theory of justice and do not engage in the well-worn debates between different variations of egalitarian justice or between egalitarian and libertarian theories. More important for my purposes is the fact that *even if* one accepts such an egalitarian

Domination." For a response from a justice-based perspective, see Arneson, "Exploitation, Domination, Competitive Markets, and Unfair Division."

57 Wertheimer, *Exploitation*.

58 For critiques of Wertheimer's view, see, e.g., Arneson, "Exploitation, Domination, Competitive Markets, and Unfair Division"; and Miklós, "Exploiting Injustice in Mutually Beneficial Market Exchange."

59 For resources, see, e.g., Rawls, *A Theory of Justice*; and Dworkin, "What Is Equality? Part 2." For needs, see, e.g., Copp, "The Right to an Adequate Standard of Living"; and Gough, "Lists and Thresholds." For capabilities, see Nussbaum, *Creating Capabilities*; and Sen, "Capability and Well-Being." For all of these theories of basic needs, resources, or capabilities, an obvious question arises about responsibility sensitiveness. A reasonable level of responsibility sensitiveness can be built into such theories, I argue in Claassen, *Capabilities in a Just Society*.

theory of justice, one can still defend a market-based conception of exploitation and hence push back against my conclusion that owners have a fiduciary duty not to exploit. For egalitarians can argue that owners of water wells cannot be made responsible for the "background injustice" of the thirsty person's lack of money. Similarly, if capitalists charge below living wages to workers, they do not exploit them since the fact that the latter have nowhere else to turn is not their fault. In both cases, it is up to the state to remedy these background injustices so that private parties can concentrate on their market-based exchanges. Exploitation can then be confined to diagnose defects in the market.

This is an important objection, so let us unpack it. State responsibility for remedying background injustices can refer to the state providing certain basic resources as public goods, either to all or targeted to those who fall below a level of basic resources through their own efforts. But many basic resources are allocated through the market as private goods—we concentrate on these here. For these private goods, egalitarian defenders of a market-based conception of exploitation could argue, the state can impose regulatory duties on owners. For example, the state could enforce maximum prices on water, just as it can enforce minimum wages. In this way, the state prevents exploitation by private owners. The objection exemplifies the standard egalitarian's position on ownership (see section 1), which grants owners' absolutist powers within the limits of the law.[60]

60 It may be confusing that I adopt an egalitarian theory of justice but place the fiduciary theory in opposition to egalitarian property theories. (See section 1.) To clarify: my fiduciary property theory does not provide a normative justification for private property itself. It can share with (and borrow from) egalitarian property theories whatever normative justification they provide for private property. I have defended my own favored egalitarian theory elsewhere (Claassen, *Capabilities in a Just Society*) but here remain agnostic. All these egalitarian theories share an argument of roughly the following form: whatever justification one gives for the property of person *A* (property contributes to the development of a person's moral powers, property provides opportunities for the exercise of a person's will, property is necessary for the protection of a person's external freedom, etc.), consistency then requires that all other persons *B*, *C*, *D*, etc. can also enjoy the condition that the justification refers to. This justification generates the egalitarian conclusion that some form of redistribution is needed (when people cannot acquire necessities through their own private activities). This form of argument is used by Murphy and Nagel in their attack on "everyday libertarianism": "The tax system … is not like an assessment of members of a department to buy a wedding gift for a colleague. It is not an incursion on a distribution of property holdings that is already presumptively legitimate. Rather, it is among the conditions that create a set of property holdings, whose legitimacy can be assessed only by evaluating the justice of the whole system, taxes included" (*The Myth of Ownership*, 36–37). Another good example is Brettschneider: "I argue that private property can be justified only in regimes in which basic material rights are guaranteed to all members of society. Specifically, private property regimes are not justifiable unless they exist in states that secure some form of a basic

My response to this objection would be that it seems to have conceded the crucial point. The state presumably regulates ownership in such cases to legally enforce a *moral* duty of owners of water wells to service thirsty persons at affordable prices. The state through its regulatory actions expresses its acceptance of this moral duty, which exists independently of the existence and actions of the state. Toy examples of desert confrontations between thirsty persons and water well owners are meant to stimulate such state-of-nature-like thinking to identify our moral duties. For owners not to fulfill these moral duties is for them to exploit nonowners. These moral duties are fiduciary since the owner has discretionary power over a vulnerable nonowner in the context of the management of a societally valuable resource.

The argument here, then, rejects a picture in which, as a matter of ideal theory, all the burdens of justice are on government. In an ideal world, owners would do their parts in realizing justice without being forced by the state's regulatory apparatus to do so. This supplements the formal account of ownership's purpose in section 3. There I argued, following Katz, that a purely formal account of ownership as office sees its purpose in *coordination*: preventing conflicts in society about who owns and therefore decides about what. Here we see that exploitation potential arises in cases where resources are critical to meeting people's justice-based claims. Then ownership acquires a second purpose: to use the resources with which owners are entrusted so as to contribute to the fulfillment of the justice-based claims of nonowners. These critical resources would include most of what is often captured under the term 'means of production', from patents and other intellectual property rights to land and real estate, natural resources and financial resources for investment.[61] To the extent that owners benefit themselves from using these resources productively to satisfy other people's needs, their personal purpose is aligned with this social purpose. But where the two conflict, then the duty not to exploit puts a strict limit on the owner's personal purposes.

Accepting the argument so far, an egalitarian property theorist might at this point argue that situations like that of the owner of the desert water well are better classified not as fiduciary duties but as *another* (nonfiduciary type of) positive duty—namely, a duty to rescue (or duty to assistance). Why not

welfare right" ("Public Justification and the Right to Private Property," 4). My fiduciary theory shares this "package deal" justificatory strategy with egalitarian property theories. One cannot justify property rights without simultaneously justifying the imposition of the relevant duties on owners. The difference is in how to understand these duties: this is embodied in the contrast between regulatory duties and fiduciary duties, which animates the main text in this section and in section 7.

61 Edmundson, "What Are 'The Means of Production'?"

endorse a property theory based on absolutist ownership, flanked with duties to rescue where necessary?

In response, I think duties of rescue differ from the fiduciary duty not to exploit in two ways. First, they are not necessarily tied to ownership but more broadly to a "capacity" to rescue, which may or may not originate in one's ownership. For example, it may also originate in being at the place of the accident, as the person walking by the pond who sees a drowning child. Second, they are not fiduciary duties, which require an exercise of discretionary power that can issue in various courses of action, but duties aimed at a more or less concrete course of action (the rescue mission). The second difference points to the fact that fiduciary relations do not arise in one-off situations where one person exercises power over another but rather in situations of "administration": where there is a structural relation of power between fiduciary and beneficiary that is institutionally embedded.[62] The administration (office) at stake for owners is *resource management*. A duty of rescue can arise also in cases where owner and nonowner are not in a relation of exploitation or not even in any kind of sustained relation, e.g., after a natural disaster (and one could accept such duties in addition to my fiduciary duties). However, when there is a situation of nonowners being structurally exploited by a practice of owners' resource management, it would be misplaced to understand the necessary duty as one where the exploiters "rescue" the persons they exploit. The fiduciary theory developed here aims to cover these situations.

Let us conclude. In this section, I have argued for the recognition of owners' fiduciary duty not to exploit nonowners. This is a moral duty. It is a further question whether this should be a *merely* moral duty, adherence to which in practice depends on the personal conscience of owners, social pressures, and protests mounted by nonowners, or whether this should also be a legally enforced duty. It is open to the egalitarian objector to concede the fiduciary moral duty in this section but to still argue that it should be legally enforced through regulation. In the next sections, I move to think about enforcement and answer this further objection.

## 5. How to Impose Legal Fiduciary Duties on Owners

Fiduciary duties can exert a moral appeal on owners. They can be the basis of protests, campaigns, media coverage, etc., in which exploited nonowners and/or other parties attribute blame to owners. But very often legal action is to be preferred over such voluntary actions, for various reasons. One is that

62 Fox-Decent, *Sovereignty's Promise*, 96–101.

legal action may be necessary given the limited societal effectiveness of moral appeals. Owners' self-interest may simply be stronger than their sense of justice. Also, people may disagree about the requirements of justice. They may not accept each other's moral judgments on the matter, and then only legal action can settle the matter in an authoritative way. Legal action over time can be seen as a "discovery procedure" for societies trying to sort out their moral intuitions about what justice requires. Finally, even when the requirements of justice are clear (in terms of the rights of those on the "demand side"), sometimes legal action is needed to determine the burden (or fair share) of each owner compared to others (the "supply side").

Legal action can take two forms in these contexts. One way is for the state to regulate situations in which owners have exploitative power over nonowners. Regulation *removes* the fiduciary situation: the discretionary power of the owner, for example, to charge anything above ten dollars per bottle of water disappears the moment the state regulates a ten-dollar maximum price. The vulnerability of the nonowner also disappears. The moral fiduciary duty in this way is converted into a legal regulatory duty. Alternatively, the moral fiduciary duty can be converted into a legal fiduciary duty. This *maintains* the fiduciary relation between owner and nonowner. In this section, I will focus on this second path, to create legal fiduciary duties. I address the issue of how to think about the choice between these two paths to legal action in section 7.

The second path itself consists of two main strategies. Effective implementation of legal fiduciary duties can be done through court enforcement or through democratization of owners' decision-making. Since the precise content of fiduciary duties needs to be tailored to specific types of ownership, I will use an example in which both strategies are actually discussed: big data companies.

Many have argued that big data companies exploit their consumers when they engage in price discrimination, fail to show products from other companies, sell the data of their consumers, etc. As an alternative to detailed state regulation that specifies which actions are forbidden, some have proposed a fiduciary alternative: these companies should be made into "information fiduciaries," imposing a fiduciary duty (also sometimes referred to as a duty of care in this context) to abstain from exploitative actions. This makes data companies themselves responsible for caring for the legitimate interests of their customers.[63] One concrete proposal is to tie the commencement of such duties to the market dominance of a company as established under antitrust law. When

63 Balkin, "Information Fiduciaries and the First Amendment"; Khan and Pozen, "A Skeptical View of Information Fiduciaries"; and Slater, "Enforcing Information Fiduciaries."

a company becomes dominant, the risk that it will exploit its consumers is material, so this would be a good point at which to introduce fiduciary duties.[64]

The scope and limits of this legal fiduciary duty would, in the final instance, have to be determined by the courts. Such an approach can hold real promise. To take another example, in 2021 a Dutch court convicted the multinational oil company Shell of breaching its duty of care by not reducing its emissions in line with the Paris Agreement on climate change.[65] The court held that human rights need to be read into the duty of care. The relevant standard was that Shell was violating the human rights of future generations to life and health. This court's judgment was made in the absence of specific legislation targeting Shell's emissions (that is, in the absence of regulatory duties). Philosophically, this can be understood as a case of structural injustice.[66]

Court enforcement is a check on an owner's (here, a company's) decisions. But some have argued that legal controls are not sufficiently effective to counter the potential for exploitation on the part of fiduciaries. Lina Khan and David Pozen sound the alarm bell over proposals to make tech platforms into information fiduciaries:

> The tension between what it would take to implement a fiduciary duty of loyalty to users, on the one hand, and these companies' economic incentives and duties to shareholders, on the other, is too deep to resolve without fundamental reform. To suggest otherwise is to risk mystification of "surveillance capitalism," entrenchment of prevailing business models, and legitimation of a wide range of troubling practices, if not also the unraveling of fiduciary law itself.[67]

In brief, Khan and Pozen believe that the incentive structures for business corporations and their shareholders are too antithetical to be remedied through the creation of fiduciary duties. Believing anything else would be naive.

64 Sauter, "A Duty of Care to Prevent Online Exploitation of Consumers?"

65 *Milieudefensie et al v. Royal Dutch Shell*, Rechtbank Den Haag [The Hague District Court], May 26, 2021, C/09/571932, ECLI:NL:RBDHA:2021:5339, available at https://uitspraken.rechtspraak.nl/#!/details?id=ECLI:NL:RBDHA:2021:5339. In the 2024 appeal case, the original verdict was overturned, although the Court of Appeal judged that breach of the duty of care is possible in principle for these kinds of cases, when the right kind of evidence is presented. See Gerechtshof Den Haag [Court of Appeal of The Hague], November 12, 2024, 200.302.332/01, ECLI:NL:GHDHA:2024:2100, available at https://uitspraken.rechtspraak.nl/details?id=ECLI:NL:GHDHA:2024:2100. The case is currently awaiting a final decision by the Dutch Supreme Court.

66 Bziuk, "Can Business Corporations Be Legally Responsible for Structural Injustice?"

67 Khan and Pozen, "A Skeptical View of Information Fiduciaries," 534–35.

Let us accept that Khan and Pozen are right about the insufficient effectiveness of legal controls of fiduciary duties in the context of the tech platforms and hence about their call for more structural reforms. Still, a fiduciary theory of property as I conceive of it has the resources to move in this more structural direction as well. This brings us to the second strategy for implementing fiduciary duties, democratization. As mentioned above, fiduciary theories are now also applied to the political relationship between citizens and governments. In democratic societies, citizens are not—at least in theory if not always in practice—passive beneficiaries of their rulers. Citizens have the right to vote their rulers in and out of office, to protest and participate in discussions about proposed legislation, etc. The fiduciary relation in a well-functioning democracy has become democratized. The control of the fiduciary's duty is then not outsourced to a third party (the courts) but enshrined in a package of democratic rights for beneficiaries.

Similarly, the relation between owners and exploited nonowners can also be democratized. This is simply one step further in terms of redressing the imbalance of power between fiduciaries and beneficiaries. It is not a denial of the fiduciary model but rather a variation within it. To illustrate, consider the economically most important property form, the corporate structure. The current situation is one in which shareholders are (*de facto* if not *de jure*) treated as the sole beneficiaries on behalf of whom corporate boards exercise their fiduciary duties. These same shareholders hold the right to vote the corporate board in and out of office. Note that this example shows the conceptual point I just made: that a fiduciary relation can simultaneously be a democratic one. Other beneficiaries (often called stakeholders in the corporate context), do not get any democratic rights and hence are often exploited. The solution, from a fiduciary perspective, can be to give democratic rights to them as well. For example, in proposals for workplace democracy, such rights are also allocated to workers.[68]

Such proposals give control rights to stakeholders, which are then shared between them and shareholders. This leaves shareholders with the economic rights to the corporation (dividends and capital gains), which often gives rise to the erroneous idea they are the "owners" of the corporation.[69] A further step would be to make one or more stakeholder groups into shareholders as well. This is effectively what happens in cooperatives and commons structures. Hence, democratic (and/or socialist) proposals for adopting such cooperative

68 Malleson, *After Occupy*; Ferreras, *Firms as Political Entities*; and Hayden and Bodie, *Reconstructing the Corporation*.

69 Ciepley, "The Anglo-American Misconception of Stockholders as 'Owners' and 'Members.'"

and common property forms are fully in the remit of the fiduciary theory of property. The theory can endorse fiduciary duties checked by courts as well as any of these democratic property forms, depending on whatever is most effective at countering exploitation. The theory does not have to pin itself on one best alternative but shows the underlying fiduciary logic common to all of them, where the variable is how radically the control rights of beneficiaries are enshrined to make sure fiduciaries do not exploit them. This institutional flexibility is a major pragmatic advantage to the fiduciary theory, while its ability to capture the normative logic underlying this full spectrum of options makes it into a generic theory of property.

## 6. DUTIES OF OWNERS IN PROPERTY THEORY

In this section, I compare the fiduciary property theory to three property theories that may seem to venture into a similar direction, since they also defend that ownership carries inherent limits (i.e., independent from concrete forms of state regulation).[70] The comparison with these theories serves to further clarify the scope and shape of the fiduciary theory.

First, Hanoch Dagan and Avihai Dorfman defend a theory in their article "Justice in Private" in which private individuals bear duties of relational justice towards each other. These duties inform different areas of private law such as contract law, tort law, and also property law. Liberal views of justice normally require individuals to recognize others only as merely formally free and equal. In Dagan and Dorfman's view, individuals must recognize each other as substantively free and equal persons in their interactions. Concretely, this requires that they "accommodate, to some extent, the personal characteristics that are necessary for the parties to recognize each other as free and equal persons."[71] This entails respecting persons' immutable circumstances and their personal choices, particularly those tied to their most fundamental "ground projects."[72] An example of a duty following from this relational theory is that landlords in

70 We might understand these theorists as elaborating on the only duty among Honoré's incidents of ownership, the duty not to harm others. See Honoré, "Ownership," 174. This duty has sometimes had a hostile reception in property theory. For example, Harris argues that such limitations are "not strictly necessary requirements of property institutions" (*Property and Justice*, 33). Another example is Penner, who argues forcefully that this duty is an unhelpful addition to Honoré's list ("The Bundle of Rights Picture of Property," 761–62).

71 Dagan and Dorfman, "Justice in Private," 190.

72 Dagan and Dorfman, "Justice in Private," 193.

their decisions to choose a tenant should not be allowed to discriminate against tenants on grounds such as religion or the color of their eyes.[73]

My theory agrees with Dagan and Dorfman's in the fundamental starting point that individuals owe each other duties of justice *qua* private individuals. It shares with them the rejection of the Rawlsian and Kantian notion of a strict division of moral labor whereby only the state is in the business of realizing justice.[74] However, Dagan and Dorfman conceive of these duties of individuals as forming a *separate* part of justice—"relational" and horizontal, in contrast to the state's effort at realizing "distributive" justice, which leads to vertical duties between state and citizens.[75] Relational justice has a *separate* task: to address the appropriate respect for personal characteristics in interpersonal relations.

By contrast, I have modelled the subject matter of the fiduciary duty on an egalitarian theory of justice. As stated in section 5, the fiduciary theory is concerned with persons' needs/capabilities for basic resources, the nonfulfillment of which marks the point at which exploitation arises. This brings concerns of distributive justice into the horizontal relation between citizens. Exploitation is cashed out in terms of this theory of egalitarian justice. The fiduciary theory is not particularly concerned with interpersonal respect for personal characteristics, as is Dagan and Dorfman's theory of relational justice. Hence, while there is a similar starting point in the project (to infuse private relations with justice requirements), the two theories have a different focus.

A second useful comparison is with Katz's theory of abuse of rights. As we saw above, Katz's property theory adheres to a formal understanding of ownership as an office. She distances herself from stewardship theories, arguing that "owners are free generally to advance their own private interests through the agendas they set for the thing (the very essence of corruption in other contexts)."[76] However, she does recognize that duties may be attached to the office of ownership via a *principle of accession*.[77] One important application of this principle of accession is Katz's argument that ownership is limited through the general private law doctrine of abuse of rights.[78] This is interesting beyond

73 Dagan and Dorfman, "Justice in Private," 175.

74 Dagan and Dorfman, "Justice in Private," 179–83.

75 Dagan and Dorfman, "Justice in Private," 174.

76 Katz, "Property's Sovereignty," 317. See also Katz, "Spite and Extortion," 1479–82; "Ownership and Offices," 274; and "It's Not Personal," 114–17.

77 Katz, "Property's Sovereignty," 306.

78 Katz, "Spite and Extortion." Another application of this principle is that governments may use owners to delegate specific public tasks to them, such as snow shoveling their part of the sidewalk so that pedestrians can safely use them. Katz refers to this phenomenon as *governing through owners*. See Katz, "Governing Through Owners."

Katz's particular account of this doctrine, since other legal scholars too may hold that the only legitimate limits on ownership rights (beyond regulatory duties) are those emanating from abuse of rights.

The doctrine of abuse of rights aims to carve out an exception to the rule that it is permissible for owners to use their property rights in ways that harm others as long as this use is not prohibited by regulation. Owners harm others in many different ways: by winning competitions, by building structures disliked by the neighbors, etc. Abuse of right obtains where the sole reason or motive of the owner's action is to impose harm on a nonowner, either for the sake of harm itself (cases of "spite") or by using their property as leverage to pressure nonowners into changing their behavior (cases of "extortion"). Katz argues that these harms can be construed as illegitimate because they violate ownership's purpose: to set an agenda for the goods in question that owners judge valuable. One example Katz cites is a case where the neighbor of musicians, tired of the sound of the musicians' daily practice, responded by himself making noise on makeshift instruments. Instead of making valuable use of these instruments, his purpose was merely to retaliate.

Abuse of right answers a normative question orthogonal to the fiduciary property theory. Applying the fiduciary theory to the case of the musicians, the question would be whether the neighbor's level of noise is such as to cause a first-order injustice to the musicians, e.g., by depriving them of their sleep (arguably a basic need). Absent such an injustice, from the point of view of the fiduciary theory, the musicians simply have to put up with the sound. It would not matter—from the point of view of diagnosing exploitation—whether the neighbor was motivated by spite or extortionary intentions. This is not incompatible with an additional normative inquiry into abuse of right, but the latter is simply a different undertaking. Katz recognizes as much when she positions the doctrine of abuse of rights in between a standpoint that gives *more* deference to owners (by accepting the agendas they set) and a standpoint that gives *less* deference to owners by requiring owners to "track the interests of others."[79] The fiduciary theory here exemplifies this latter standpoint, searching for the conditions under which owners must indeed track the interests of others.

Finally, Gregory Alexander's progressive property theory is important as a point of comparison. Alexander's theory does go into an interest-tracking direction. He proposes that owners are subject to a social obligation norm.[80] Alexander's theory, like the fiduciary theory defended here, argues that owners

79 Katz, "Spite and Extortion," 1472.

80 Alexander, "The Social-Obligation Norm in American Property Law" and *Property and Human Flourishing*.

have obligations to nonowners that are "conceptually entailed in ownership."[81] Alexander grounds these duties in a "human flourishing theory" that takes its inspiration from Amartya Sen and Martha Nussbaum. He singles out four capabilities as essential for human flourishing: capabilities to life, freedom, practical reasoning, and sociability.[82] Finally, his theory is built on a concept of community, since (following Charles Taylor's *social thesis*) individuals can flourish only in communities and hence have a duty to support the existence of these communities.[83] With respect to ownership, this leads Alexander to endorse the general principle that "the social obligation may require the owner to provide resources, in ways that are appropriate to that owner, to others in the owner's community (or communities) where necessary to support the development of their requisite capabilities."[84]

The similarities between Alexander's theory and the fiduciary theory may seem strong, as both of them widen the concept of an owner's duties. However, there are also striking differences. First, the fiduciary theory is not wedded to an understanding of ownership duties that is tied to reciprocating towards other members of a community. Instead, fiduciary duties are triggered whenever owners have the power to exploit the lives of nonowners, whether or not the nonowners are part of the same community. Second and most importantly, Alexander restricts the scope of ownership's duties through a further distinction between *general* and *special* obligations. He argues the former are about "basic material conditions that a humane liberal society needs for its members to flourish"; they are owed to society in general by everyone and are redistributive, paid through taxation.[85] Special duties are owed by owners to their particular communities. They are limited to particular incidents of ownership, and nonredistributive.[86] This distinction allows Alexander to focus attention on the latter, and hence he at length discusses particular duties of owners, e.g., in the context of land reparations, access to beaches, and preservation of historic buildings.

The fiduciary theory presented here, by contrast, does not accept this distinction between general and special obligations. I have argued that when owners control resources that are critical to nonowners, they have opportunities to exploit the latter. This can sometimes be remedied through taxation, but

81 Alexander, *Property and Human Flourishing*, xv.

82 Alexander, *Property and Human Flourishing*, 7.

83 Alexander, *Property and Human Flourishing*, 44–55.

84 Alexander, *Property and Human Flourishing*, 60.

85 Alexander, *Property and Human Flourishing*, 56.

86 Alexander, *Property and Human Flourishing*, 56–62.

on other occasions, regulation or legal fiduciary duties are a better response. (See also the next section.) As a consequence, Alexander's distinction allows him to contain owners' inherent duties to a relatively small set of particular situations, but if I am right, a larger set of situations related to economically vital resources can be classified as giving owners exploitative power, triggering fiduciary duties.

## 7. THE NORMATIVE UPSHOT: RECONCEIVING THE RELATION BETWEEN STATES AND PRIVATE OWNERS

As we saw in section 1, on an absolutist understanding of ownership, the state can impose regulatory duties to restrict the power of owners. Absolutist ownership is always exercised within the limits of the law, and an egalitarian state can set strict limits. The picture emerging from this is a dualist one, where owners concentrate on their own private interests (however conceived), and states provide a counterweight through regulation. In this section, I want to compare this picture with the fiduciary property theory developed over the course of this paper.[87]

The dualist picture has been forcefully defended by Arthur Ripstein in his critical discussion of stewardship theories of property. Imagining someone who argues that property needs to be rethought in light of challenges such as climate change, he replies:

> It is incumbent on sovereigns to preserve the natural conditions of the continued existence of their societies, and so, in the service of that mandatory purpose, to constrain and coordinate the ways in which land is used and other resources depleted within their political societies, to restrict deforestation or impose carbon taxes, and so on. But these are fundamentally public matters, not because private owners should take no moral interest in them, but rather because any solution to them is essentially public and global, not a matter of the state reminding particular individual citizens and owners about the specific things that they were already under an obligation to do in their capacity as owners of land or other property.[88]

With this, Ripstein reaffirms the traditional division of moral labor between states and owners. For owners, a quasi-mechanical obedience to these lawfully

87 Elsewhere I elaborate this comparison in terms of two alternative views of representation: one indirect (via the state), the other direct (at the level of ownership). See Claassen, "Property as Power."

88 Ripstein, "Property and Sovereignty," 268.

imposed duties is enough. The traction of this objection goes beyond Ripstein's own Kantian property theory. A similar objection can be made using a variety of normative bases. Earlier, in section 1, we saw how libertarian and egalitarian theories both are wedded to an absolutist conception of property, even if they disagree about how many regulatory limits to impose. How to respond?

Ripstein's reference to being "under an obligation" most probably refers to moral obligations. (If we understand them as legal obligations, his claim is trivially true: before the state's declaration of a law, no legal obligations exist.) I argue the claim is untenable. Even before the state has decided that the use of child labor in the production of textiles is exploitative and should be prohibited, it already is exploitative. Even before the state has decided that making consumers addicted to nicotine is exploitative, it already is so. If there is a prepolitical, moral claim (from the point of view of a person "in the state of nature") to see one's *de facto* possessions recognized as one's legitimate property under law, then there is also a prepolitical, moral claim that owners should not exploit nonowners.[89] This amounts to the recognition of a moral fiduciary duty, on the level of fiduciary *ethics*.

Of course, in less clear-cut cases, a legal determination is necessary to establish what is and what is not going to count as exploitative; and legal enforcement is certainly necessary to make this effective in practice. But at the political level, Ripstein excludes without argument fiduciary duties as potential solutions that a lawgiver might adopt (and often does adopt). Of course, lawmakers also regularly adopt Ripstein's favored nonfiduciary solution, a regime of regulatory duties. From a claim of exploitation in fiduciary ethics, we can still go both ways at the level of legal action. It all depends on the relative effectiveness of external regulation vis-à-vis fiduciary governance, which is hard to decide *a priori*. But conversely, there is no reason to presuppose that fiduciary governance can *never* be more effective.

An analogy with multiple layers of public governance is helpful.[90] States make national laws to govern the internal affairs of their citizens. However, this leaves many matters unresolved. Some are left to lower-level governments, like provinces and municipalities. These do not mechanistically apply national law but form a separate sphere of jurisdiction. Different layers of authority are nested into each other. Similarly, fiduciary ownership is decentralized governance, under the larger umbrella of the nation-state. This decentralization empowers fiduciary owners themselves, but this power can be checked at the

89 If we, along Kantian lines, call the former *provisionally rightful possession* (see Kant, *The Metaphysics of Morals*, 410 [6:257]), then let us call the latter a *provisional duty to nonexploitation*.

90 See González-Ricoy, "Little Republics," 115.

level of courts, which decide on a case-to-case basis but simultaneously develop a legal set of precedents over time. Moreover, as we have seen, it can also be checked through more democratic arrangements at the level of ownership.

There can be good reasons for choosing this fiduciary route. For one thing, the same information problems that make it difficult to make a complete contract between private parties (owners and nonowners) may also make it equally (or even more) difficult for government as an external party to regulate. Also, regulation can be ineffective not because of informational problems but because it is introduced in situations where interests are opposed. Regulatory efforts often meet with resistance, either overt or covert, and powerful owners try to influence the political system to have favorable regulations applied. Regulation of property has achieved impressive results in the "regulatory state" that has been built up since the Second World War in many countries.[91] Still, as Paul Babie argues, "recognizing the place of regulation within the concept of private property is not to say that a system achieves perfect symmetry. Comparing the legal protection of choice to regulation always yields a surplus of individual choice and a deficit of regulation."[92] The reason for this is ultimately that the private/public dualism that Ripstein celebrates puts owners in a position where they do not have to take wider interests into account.

Maintaining an absolutist preference for state regulation ultimately means outsourcing the problem of ownership's exploitation potential to states with limited regulatory capacities. The fiduciary perspective opens up a route that would change the mindset of owners by requiring them to adopt a broader stance. One author, in the course of defending the proposal mentioned earlier to turn platform providers into information fiduciaries, sums up the case for this perspective when she writes:

> The fiduciary model is a happy marriage of pragmatism and aspirational ethics. Surely, not every doctor, lawyer, or other fiduciary professional is a paragon of virtue or is able to provide the best advice for their clients in every case. But for the most part, many of us do in fact trust the fiduciaries in our lives, and when they let us down, remedies are available.[93]

If this would work for owners, it would lead to a fundamental shift in our economic life, especially in the current context in which owners of productive assets too often exploit the earth's resources and transgress planetary boundaries, lobby for minimal regulatory burdens, and subvert regulation that has

91 Moran, "Understanding the Regulatory State."

92 Babie, "Private Property Suffuses Life," 139.

93 Slater, "Enforcing Information Fiduciaries," 89.

been passed into law. The alternative would be a relation whereby states and private owners are not antagonists in a cat-and-mouse regulation game but rather cooperating partners who work together toward a less exploitative economy.

## 8. Conclusion

Ideas about owners as stewards have an old pedigree. As we have seen, religious and Indigenous traditions have been protagonists of such views, as have recent environmentalist theories with respect to natural resources. However, beyond these particular contexts, no secular and general theory conceiving of ownership as stewardship has been developed. An absolutist understanding of property underlies both libertarian and egalitarian property theories.

In this paper, I have used the technical notion of a fiduciary relation to propose such a theory. Fiduciary relations are recognized in private law (albeit not for ownership) and increasingly used to understand public office as well. First, the theory proposed here is grounded in Katz's legal theory of ownership as a state-mandated office. Second, it makes use of the economic analysis of incomplete contracting and externalities, in combination with an ethics of exploitation, to diagnose when nonowners deserve protection through moral fiduciary duties. Third, it argues these moral duties can be discharged in various ways, through regulatory duties enforced by the state, as well as through legal fiduciary duties checked by courts or various democratic forms of ownership.

The central point is that the granting of ownership rights and the imposition of fiduciary duties should not be separated. Property rights confer powers on owners to control assets—powers that, by virtue of their social recognition, enable owners to exercise power over nonowners, which can turn into exploitation. Legitimation of such powers (if at all) often requires the recognition of fiduciary duties on owners to abstain from exploitation. Both are a package deal. To externalize the question of duties and define the office of owners only through their rights would be similar to defining the position of political rulers through their prerogatives alone, failing to describe their duties to govern for their people. While the cases are not exactly similar—owners, unlike politicians, retain an allowance to act for self-serving purposes within these limits—the analogy is stronger than often acknowledged.

If this is convincing, then the marginal existence of stewardship views of property is unwarranted. They are not tied to particular religious beliefs that ground an owner's assignment to manage all earthly resources in a divine command. Instead, stewardship duties can be grounded in mainstream egalitarian theories of distributive justice. They are not tied to a particular set of resources, such as in theories of environmental ethics that focus on natural

resources. Instead, stewardship duties can extend to whatever resources the theory of justice recognizes as vital for all persons to sustain their basic needs or capabilities.[94]

*Utrecht University*
*r.j.g.claassen@uu.nl*

## REFERENCES

Alexander, Gregory. *Property and Human Flourishing*. Oxford University Press, 2018.

———. "The Social-Obligation Norm in American Property Law." *Cornell Law Review* 94, no. 4 (2009): 745–819.

Aquinas, Thomas. *Political Writings*. Edited by R. W. Dyson. Cambridge University Press, 2002.

Aristotle. "Politics." In *The Complete Works of Aristotle: The Revised Oxford Translation*, vol. 2, edited by Jonathan Barnes. Princeton University Press, 1984.

Arneson, Richard. "Exploitation, Domination, Competitive Markets, and Unfair Division." *Southern Journal of Philosophy* 54, no. 1 (2016): 9–30.

Attfield, Robin. *The Ethics of the Global Environment*. Edinburgh University Press, 2015.

Babie, Paul. "Private Property Suffuses Life." *Sydney Law Review* 39, no. 1 (2017): 135–46.

Balkin, Jack M. "Information Fiduciaries and the First Amendment." *UC Davis Law Review* 49, no. 4 (2016) 1183–234.

Becker, Lawrence. *Property Rights: Philosophic Foundations*. Routledge, 1977.

Blackstone, William. *Commentaries on the Laws of England*. Bk. 2, *Of the Rights of Things*, edited by Simon Stern. Oxford University Press, 2016.

Brettschneider, Corey. "Public Justification and the Right to Private Property: Welfare Rights as Compensation for Exclusion." In *Property-Owning*

94 I would like to thank audiences at Stockholm University and Utrecht University. For feedback on (much) earlier drafts of this text, I thank Bas van Bavel, Michael Bennett, Friedemann Bieber, Barbara Bziuk, John Christman, Hanoch Dagan, Liesbeth Feikema, Thomas Fossen, Ingrid, Robeyns, Larissa Katz, Tully Rector, Yara al Salman, Philipp Stehr, Eva Vermeulen, and Stuart Write. Finally, the reviewers for this journal I thank for their very helpful comments on earlier versions of this paper, which have very much improved the paper. I acknowledge funding under the research programme Private Property and Political Power, grant number 360-20-390, financed by the Dutch Research Council (NWO).

*Democracy: Rawls and Beyond*. Wiley-Blackwell, 2012.
Bziuk, Barbara. "Can Business Corporations Be Legally Responsible for Structural Injustice? The Social Connection Model in (Legal) Practice." *Critical Review of International Social and Political Philosophy* (forthcoming). Published online ahead of print, September 7, 2022. https://doi.org/10.1080/13698230.2022.2120732.
Carpenter, Kristen A., Sonia K. Katyal, and Angela R. Riley. "In Defense of Property." *Yale Law Journal* 118, no. 6 (2009): 1022–125.
Cassasas, David, and Jordi Mundó. "Property as a Fiduciary Relationship and the Extension of Economic Democracy: What Role for Unconditional Basic Income?" *Theoria* 69, no. 171 (2022): 74–96.
Ciepley, David. "The Anglo-American Misconception of Stockholders as 'Owners' and 'Members': Its Origins and Consequences." *Journal of Institutional Economics* 16, no. 5 (2020): 623–42.
Claassen, Rutger. *Capabilities in a Just Society. A Theory of Navigational Agency*. Cambridge University Press, 2018.
———. "Property and Political Power: Neo-Feudal Entanglements." In *Positive Liberty: Past, Present and Future*, edited by John Christman. Cambridge University Press, 2021.
———. "Property as Power: A Theory of Representation." *Journal of Social Philosophy* (forthcoming). Published online ahead of print, August 27, 2024. https://doi.org/10.1111/josp.12587.
Claassen, Rutger, and Larissa Katz. "Property: Authority Without Office?" *Journal of Law and Political Economy* 3, no. 3 (2023): 570–75.
Coase, R. H. "The Problem of Social Cost." *Journal of Law and Economics* 3 (1960): 1–44.
Cohen, Morris. "Property and Sovereignty." *Cornell Law Review* 13, no. 1 (1927): 8–30.
Copp, David. "The Right to an Adequate Standard of Living: Justice, Autonomy, and the Basic Needs." *Social Philosophy and Policy* 9, no. 1 (1992): 231–61.
Criddle, Evan. "Stakeholder Fiduciaries." In *Fiduciairies and Trust: Ethics, Politics, Economics and Law*, edited by Paul Miller and Matthew Harding. Cambridge University Press, 2020.
Criddle, Evan, Evan Fox-Decent, Andrew Gold, Sung Hui Kim, and Paul Miller, eds. *Fiduciary Government*. Cambridge University Press, 2018.
Dagan, Hanoch. *A Liberal Theory of Property*. Cambridge University Press, 2021.
———. *Unjust Enrichtment: A Study of Private Law and Public Values*. Cambridge University Press, 1997.
Dagan, Hanoch, and Avihay Dorfman. "Justice in Private: Beyond the Rawlsian Framework." *Law and Philosophy* 37, no. 2 (2018): 171–201.

Dagan, Hanoch, and Elizabeth Scott. "Reinterpreting the Status-Contract Divide: The Case of Fiduciaries." In *Contract, Status, and Fiduciary Law*, edited by Paul Miller and Andrew Gold. Oxford University Press, 2016.

Demsetz, Harold. "Toward a Theory of Property Rights." *American Economic Review* 57, no. 2 (1967): 347–59.

di Robilant, Anna. *The Making of Modern Property*. Cambridge University Press, 2023.

Domènech, Antoni, and María Julia Bertomeu. "Property, Freedom and Money: Modern Capitalism Reassessed." *European Journal of Social Theory* 19, no. 2 (2016): 245–63.

Dworkin, Ronald. "What Is Equality? Part 2: Equality of Resources." *Philosophy and Public Affairs* 10, no. 4 (1981): 283–345.

Edmundson, William. "What Are 'the Means of Production'?" *Journal of Political Philosophy* 28, no. 4 (2020): 421–37.

Essert, Christopher. "The Office of Ownership." *University of Toronto Law Review* 63, no. 3 (2013): 418–61.

———. "The Office of Ownership Revisited." *University of Toronto Law Review* 70, no. 2 (2020): 287–98.

Ferreras, Isabelle. *Firms as Political Entities: Saving Democracy Through Economic Bicameralism*. Cambridge University Press, 2017.

Finnis, John. *Aquinas: Moral, Political, and Legal Theory*. Oxford University Press, 1998.

Fox-Decent, Evan. "Challenges to Public Fiduciary Theory: An Assessment." In *Research Handbook on Fiduciary Law*, edited by D. Gordon Smith and Andrew S. Gold. Edward Elgar, 2018.

———. *Sovereignty's Promise: The State as Fiduciary*. Oxford University Press, 2011.

Frank, Jill. *A Democracy of Distinction: Aristotle and the Work of Politics*. University of Chicago Press, 2005.

Frankel, Tamar. *Fiduciary Law*. Oxford University Press, 2011.

Galoob, Stephen R., and Ethan J. Leib. "Intentions, Compliance, and Fiduciary Obligations." *Legal Theory* 20, no. 2 (2014): 106–32.

Garnsey, Peter. *Thinking About Property: From Antiquity to the Age of Revolution*. Cambridge University Press, 2007.

Gaus, Gerald, and Loren Lomasky. "Are Property Rights Problematic?" *Monist* 73, no. 4 (1990): 486–503.

González-Ricoy, Iñigo. "Little Republics: Authority and the Political Nature of the Firm." *Philosophy and Public Affairs* 50, no. 1 (2022): 90–120.

Gough, Ian. "Lists and Thresholds: Comparing the Doyal-Gough Theory of Human Need with Nussbaum's Capabilities Approach." In *Capabilities,*

*Gender, Equality: Towards Fundamental Entitlements*, edited by Flavio Comim and Martha Nussbaum. Cambridge University Press, 2014.

Grey, Thomas. "The Disintegration of Property." In *Nomos, Volume 22: Property*, edited by J. Roland Pennock and John W. Chapman. New York University Press, 1980.

Harris, J. W. *Property and Justice*. Clarendon Press, 1996.

Hausman, Daniel. "When Jack and Jill Make a Deal." *Social Philosophy and Policy* 9, no. 1 (1992): 95–113.

Hayden, Grant, and Matthew Bodie. *Reconstructing the Corporation: From Shareholder Primacy to Shared Governance*. Cambridge University Press, 2020.

Hegel, G. W. F. *Elements of the Philosophy of Right*. Cambridge University Press, 1991.

Honoré, A. M. "Ownership." In *Making Law Bind: Essays Legal and Philosophical*. Clarendon Press, 1987.

Ireland, Paddy. *Property in Contemporary Capitalism*. Bristol University Press, 2024.

Kant, Immanuel. "The Metaphysics of Morals." In *Practical Philosophy*, edited by Paul Guyer. Cambridge Edition of the Works of Immanuel Kant. Cambridge University Press, 1996.

Katz, Larissa. "Exclusion and Exclusivity in Property Law." *University of Toronto Law Journal* 58, no. 3 (2008): 275–315.

———. "Governing Through Owners: How and Why Formal Private Property Rights Enhance State Power." *University of Pennsylvania Law Review* 160, no. 7 (2012): 2029–60.

———. "It's Not Personal: Social Obligations in the Office of Ownership." *Cornell Journal of Law and Public Policy* 29, no. 10 (2020): 101–19.

———. "Ownership and Offices: The Building Blocks of the Legal Order." *University of Toronto Law Review* 70, no. 2 (2020): 267–86.

———. "Property's Sovereignty." *Theoretical Inquiries in Law* 18, no. 2 (2017): 299–328.

———. "Spite and Extortion: A Jurisdictional Principle of Abuse of Property Right." *Yale Law Journal* 122, no. 6 (2013): 1444–82.

Khan, Lina M., and David E. Pozen. "A Skeptical View of Information Fiduciaries." *Harvard Law Review* 133, no. 2 (2019): 497–541.

Lametti, David. "The Concept of Property: Relations Through Objects of Social Wealth." *University of Toronto Law Journal* 53, no. 4 (2003): 325–78.

———. "The Objects of Virtue." In *Property and Community*, edited by Gregory Alexander and Eduardo Penalver. Oxford University Press, 2010.

Lees, E. "Property in the Anthropocene." *William and Mary Environmental Law*

*and Policy Review* 43 (2019): 541–93.
Leib, Ethan J., David L. Ponet, and Michael Serota. "Mapping Public Fiduciary Relationships." In *Philosophical Foundations of Fiduciary Law*, edited by Andrew S. Gold and Paul B. Miller. Oxford University Press, 2014.
Locke, John. *Two Treatises of Government*. Cambridge University Press, 2003.
Malleson, Tom. *After Occupy: Economic Democracy for the 21st Century*. Oxford University Press, 2014.
Markovits, Daniel. "Sharing *Ex Ante* and Sharing *Ex Post*." In *Philosophical Foundations of Fiduciary Law*, edited by Andrew S. Gold and Paul B. Miller. Oxford University Press, 2014.
Miklós, András. "Exploiting Injustice in Mutually Beneficial Market Exchange: The Case of Sweatshop Labor." *Journal of Business Ethics* 156, no. 1 (2019): 59–69.
Miller, Paul. "The Fiduciary Relationship." In *Philosophical Foundations of Fiduciary Law*, edited by Andrew S. Gold and Paul B. Miller. Oxford University Press, 2014.
Miller, Paul, and Andrew Gold. "Fiduciary Governance." *William and Mary Law Review* 57, no. 2 (2015): 513–86.
Moran, Michael. "Understanding the Regulatory State." *British Journal of Political Science* 32, no. 2 (2002): 391–413.
Murphy, Liam, and Thomas Nagel. *The Myth of Ownership: Taxes and Justice*. Oxford University Press, 2002.
Mussell, Helen J. "Theorising the Fiduciary: Ontology and Ethics." *Journal of Business Ethics* 186, no. 2 (2023) 293–307.
Narveson, Jan. *The Libertarian Idea*. Broadview Press, 2001.
Nozick, Robert. *Anarchy, State, and Utopia*. Blackwell Publishing, 1974.
Nussbaum, Martha. *Creating Capabilities: The Human Development Approach*. Belknap Press, 2011.
Olsen, Erik. "The Early Modern 'Creation' of Property and Its Enduring Influence." *European Journal of Political Theory* 21, no. 1 (2021): 111–33.
Penner, James E. "The Bundle-of-Rights Picture of Property." *UCLA Law Review* 43, no. 3 (1996): 711–820.
———. "Private Law Offices." *University of Toronto Law Review* 70, no. 2 (2020): 299–313.
Pierson, Christopher. *Just Property: A History in the Latin West*. 3 vols. Oxford University Press, 2013–20.
Pontifical Council for Justice and Peace. *Compendium of the Social Doctrine of the Church*. Libreria Editrice Vaticana, 2022.
Radin, Margaret. "Property and Personhood." *Stanford Law Review* 34, no. 5 (1982): 957–1015.

Rawls, John. *A Theory of Justice*. Rev. ed. Harvard University Press, 1999.
Ripstein, Arthur. *Force and Freedom: Kant's Legal and Political Philosophy*. Harvard University Press, 2009.
———. "Property and Sovereignty: How to Tell the Difference." *Theoretical Inquiries in Law* 18, no. 2 (2017): 243–68.
Robé, Jean-Philippe. *Property, Power and Politics: Why We Need to Rethink the World Power System*. Bristol University Press, 2020.
Rodgers, Christopher. "Nature's Place? Property Rights, Property Rules and Environmental Stewardship." *Cambridge Law Journal* 68, no. 3 (2009): 550–74.
Rose, Carol M. "Property as Wealth, Property as Propriety." *Nomos* 33 (1991): 223–47.
Ryan, Alan. *Property and Political Theory*. Basil Blackwell, 1984.
Sait, Siraj, and Hilary Lim. *Land, Law, and Islam: Property and Human Rights in the Muslim World*. Zed Books, 2006.
Samet, Iris. *Equity: Conscience Goes to Market*. Oxford University Press, 2018.
Sauter, Wolf. "A Duty of Care to Prevent Online Exploitation of Consumers? Digital Dominance and Special Responsibility in EU Competition Law." *Journal of Antitrust Enforcement* 8, no. 2 (2020): 406–27. Correction published in *Journal of Antitrust Enforcement* 8, no. 3 (2020): 649.
Scott, Robert E. "The Law and Economics of Incomplete Contracts." *Annual Review of Law and Social Science* 2 (2006): 279–97.
Sen, Amartya. "Capability and Well-Being." In *The Quality of Life*, edited by Martha Nussbaum and Amartya Sen. Clarendon Press, 1993.
Singer, Joseph. *The Edges of the Field: Lessons on the Obligations of Ownership*. Beacon Press, 2001.
———. "How Property Norms Construct the Externalities of Ownership." In *Property and Community*, edited by Gregory Alexander and Eduardo Peñalver. Oxford University Press, 2010.
Slater, Felicity. "Enforcing Information Fiduciaries." *Boston University Journal of Science and Technology Law* 28, no. 1 (2022): 92–134.
Smith, Lionel. "Fiduciary Relationships: Ensuring the Loyal Exercise of Judgement on Behalf of Another." *Law Quarterly Review* 130 (2014): 608–34.
———. "Parenthood Is a Fiduciairy Relationship." *University of Toronto Law Journal* 70, no. 4 (2020): 395–452.
Tierney, Brian. *The Idea of Natural Rights: Studies on Natural Rights, Natural Law and Church Law, 1150–1625*. Scholars Press, 1997.
Tomasi, John. *Free Market Fairness*. Princeton University Press, 2012.
Viehoff, Daniel. "Legitimate Injustice and Acting for Others." *Philosophy and Public Affairs* 50, no. 3 (2022): 301–74.

Vrousalis, Nicholas. "Exploitation, Vulnerability, and Social Domination." *Philosophy and Public Affairs* 41, no. 2 (2013): 131–57.
Waldron, Jeremy. *The Right to Private Property*. Clarendon Press, 1988.
Wertheimer, Alan. *Exploitation*. Princeton University Press, 1996.
Zingales, Luigi. "Corporate Governance." In *The New Palgrave Dictionary of Economics and the Law*, edited by Peter Newman. Palgrave Macmillan, 1998.

*Journal of Ethics and Social Philosophy*
VOL. 30, NO. 2 · APRIL 2025
*https://doi.org/10.26556/jesp.v30i2.4043*
© 2025 Author

# MORAL DISAGREEMENT AND THE QUESTION UNDER DISCUSSION

*Stina Björkholm*

When conservatives and progressives discuss the moral status of abortion, they can disagree despite their different views of what it is for an act to be wrong or right. Contextualists about normative language have famously been challenged to give an account of what speakers disagree about in such cases.[1] They maintain that the extensions of moral expressions are determined by an agent-sensitive parameter at the context of utterance. If speakers refer to different properties when they use moral expressions, they turn out to be talking past one another when they make seemingly conflicting moral claims. Hence, the conservative and the progressive seem unable to have a first-order moral disagreement about whether abortion is wrong, if contextualism is correct.

This paper addresses how contextualists might explain moral disagreement by drawing attention to the broader conversational context in which utterances are made rather than trying to locate a conflict between the semantic contents expressed or contents that are pragmatically conveyed by utterances.[2] Such accounts of disagreement focus on the shared assumptions among the interlocutors about the background of their communicative exchange. I defend an account according to which the shared conversational background, *inter alia*,

1 This problem has been widely discussed in relation to contextualism (and relativism) about taste predicates, epistemic modals, and other normative and evaluative expressions. See, for example, Kölbel, "Faultless Disagreement"; MacFarlane, "Relativism and Disagreement"; Huvenes, "Varieties of Disagreement and Predicates of Taste" and "Disagreement Without Error"; Cohnitz and Marques, "Disagreements"; Marques, "Doxastic Disagreement"; Brendel, "Contextualism, Relativism, and the Problem of Lost Disagreement"; Khoo, "The Disagreement Challenge to Contextualism"; and Zeman, "Faultless Disagreement." I focus on moral disagreement in this paper. However, there is potential for the positive account presented here to be further developed to explain disagreement in these other areas as well.

2 Plunkett and Sundell, "Disagreement and the Semantics of Normative and Evaluative Terms" and "Metalinguistic Negotiation and Speaker Error"; Pérez Carballo and Santorio, "Communication for Expressivists"; and Khoo and Knobe, "Moral Disagreement and Moral Semantics."

includes questions that the interlocutors mutually aim to resolve.[3] In a nutshell, the proposal is that speakers can use moral expressions with different extensions but still mutually accept that they have a shared question that they aim to resolve.

The structure of the paper is as follows. I start in section 1 by clarifying the way that the problem of lost disagreement is understood in this paper and provide reason to pursue accounts of disagreement that focus on the broader conversational setting. In section 2, I present problems for three such accounts. In section 3, I present my own preferred view and explain how it avoids the problems from section 2. Lastly, section 4 concludes.

## 1. THE PROBLEM OF LOST DISAGREEMENT

Contextualists maintain that a speaker's personal moral outlook, or the standards or norms of her society, is relevant to determine the extension of a moral expression at a context of utterance. If the extension of a moral expression varies depending on the context of utterance, two speakers who embrace different moral norms or come from different societies will refer to different properties when they use that expression. For instance, consider the following two utterances:

1. "Abortion is wrong."
2. "Abortion is not wrong."

If the speaker of 1 accepts a moral norm according to which an act is wrong iff it has the property (or a set of properties) of being *F*, and the speaker of 2 accepts another moral norm according to which an act is wrong iff it is *G*, they will refer to different properties when they use 'wrong'. The speakers are therefore *talking past* one another, since one talks about whether abortion is *F*, and the other talks about whether it is *G*. *They are not talking about the same thing*. We might, for instance, think of a disagreement between a conservative and a progressive who assert 1 and 2 as being of this kind. While this problem also arises for other normative disagreements in which speakers assume different norms (such as rationality or taste), I will focus on moral disagreement here.

The main challenge for contextualists will not be taken to be that they must show that the contents of 1 and 2 are *exclusionary,* i.e., that both claims cannot be true.[4] The reason is that it is controversial whether moral disagreements

3 Roberts, "Context in Dynamic Interpretation" and "Information Structure in Discourse."

4 The claim that moral disagreements are exclusionary is put forward in Streumer, *Unbelievable Errors*, for instance.

between people who embrace different norms or standards are intuitively exclusionary. Studies suggest that people are more inclined to think that at least some moral disagreements are exclusionary, but they do not treat moral disagreements on a par with disagreements about matters of fact.[5] Moreover, people are more inclined to think that disagreements between people within the same culture are exclusionary than disagreements between people from different cultures—and even less so for potential disagreements between humans and extraterrestrials.[6] These studies might not vindicate that truth in moral matters is relative, but they provide enough evidence that exclusion is not to be treated as a desideratum for a satisfactory account of moral disagreement.

Instead, I consider the main challenge for contextualism to be that intuitively, the speakers are not talking past each other; they share a common topic and disagree about it. The problem for contextualism is that this intuition constitutes evidence that moral expressions should not be given a contextualist semantics since other context-sensitive expressions do not render the same intuitions. In cases of standard context sensitive expressions such as 'I', 'here', and 'now', there is not even a seeming disagreement. If two different speakers assert "I am tired" and "I am not tired," it is obvious that they do not disagree. In other cases of context-sensitive expressions, the speakers might appear to disagree, but once they clarify their terminology, they will realize that the disagreement is *merely verbal*. For instance, when two speakers assert sentences such as "Jim is tall" and "Jim is not tall" and intend different contrast classes for tallness (e.g., tall *for a twelve-year old boy* and not tall *for a grown-up man*), their disagreement will resolve once this difference is brought to light. The issue for contextualism is that unlike these cases, there is intuitively a disagreement in 1 and 2, and the disagreement is not resolved once the speakers' divergent moral norms are clarified. This constitutes evidence that moral expressions are not context sensitive and therefore risk undermining the view.[7]

5 Goodwin and Darley, "The Psychology of Meta-Ethics," "The Perceived Objectivity of Ethical Beliefs," and "Why Are Some Moral Beliefs Perceived to Be More Objective than Others?"; Beebe, "Moral Relativism in Context"; Wright et al., "The Meta-Ethical Grounding of Our Moral Beliefs"; Sarkissian, "Aspects of Folk Morality"; Pölzler, "Revisiting Folk Moral Realism"; Pölzler and Wright, "Anti-Realist Pluralism" and "An Empirical Argument Against Moral Non-Cognitivism."

6 Sarkissian, "Aspects of Folk Morality."

7 There is a version of this kind of argument that can be pressed not only for contextualists but also for naturalists who maintain that all speakers invariably refer to the same property when they use the expression 'wrong'. The problem is put forward by Eklund, who argues that even if our concept 'wrong' invariably refers to the same property regardless of context of utterance, there might be another society or language that includes a 'wrong'-*like* concept, 'wrong*', which invariantly refers to another property ("Alternative Normative

While I focus here on the moral disagreement between the speakers of 1 and 2, the problem is more general, since it arises for other forms of normative and evaluative language. Moreover, the need for an account of common topic for expressions that are not coextensive has even turned out to generalize beyond the area of normative and evaluative language.[8] There are many cases in which people have (or have had) very different uses of an expression, but disagreements involving this expression are not easily resolved by simply clarifying one's terminology. Examples include not only normative expressions such as 'wrong' and 'right' but also expressions such as 'woman', 'know', 'person', and 'atom'. If some of these other expressions are also context dependent or polysemous, then the fact that disagreement persists in the moral case need not provide conclusive evidence that the extensions of moral predicates cannot be context sensitive. Rather, intuitions can vary between different cases in which speakers use an expression with different extensions. There are cases in which there is no intuition of disagreement at all (e.g., with 'I'), cases in which there is an intuition of disagreement that is defeated once it is acknowledged to be merely verbal (e.g., with 'tall'), and cases in which the disagreement intuition persists even if a divergent terminology is revealed (e.g., with 'wrong', 'woman', 'atom', etc.).

This suggests that speakers may share a *common topic* that does not correspond to common extension. As Herman Cappelen puts it, "Sameness of topic doesn't track sameness of extension."[9] Still, contextualists owe an account of how to accommodate that speakers talk about the same thing, even if they use an expression with different extensions. What is normative disagreement without coextension, and how does it differ from merely verbal disagreements?

Before I move on to consider the approach to disagreement that I focus on here, I will briefly discuss some of the other proposed solutions that have been defended.[10] One option is to say that it is not a speaker's individual norm that determines the extension of the moral term but instead the norm accepted by the speaker's whole community or society. If so, disagreements between speakers of the same community constitute a disagreement in belief over whether an

Concepts" and *Choosing Normative Concepts*). For reasons of space, I put aside this problem here, but I believe that a solution that I put forward in section 3.1 below might be able to solve this problem too.

8 The focus on common topic in relation to these concepts comes from the field of conceptual engineering. For overviews, see Cappelen, *Fixing Language*; and Cappelen et al., *Conceptual Engineering and Conceptual Ethics*.

9 Cappelen, *Fixing Language*, 110.

10 For helpful overviews of the problem of disagreement, see Cohnitz and Marques, "Disagreements"; Tersman's entries for "moral disagreement" in the *International Encyclopedia of Ethics* and the *Stanford Encyclopedia of Philosophy*; and Björnsson, "The Significance of Ethical Disagreement for Theories of Ethical Thought and Talk."

act satisfies this common norm.[11] This response does not, however, account for disagreements between speakers from different communities.[12]

Another option is to explain moral disagreement as a disagreement in attitude and to maintain that the relevant attitude is pragmatically expressed by the utterance of a moral sentence.[13] The main problem for such accounts is to explain what it is for two attitudes to conflict in a way that constitutes a disagreement. One proposal is that two attitudes conflict when an individual cannot rationally accept both attitudes at the same time. This is criticized by Teresa Marques, who argues that this account of conflict is unsatisfactory because two mental states may give rise to intrapersonal conflict without thereby giving rise to interpersonal conflict.[14] For example, there is a conflict between the intention to stop smoking and the intention to have a cigarette when these intentions are held by the same person, but not if they are held by different people. Moreover, Marques further argues that it is controversial whether combinations of noncognitive attitudes can be irrational at all.[15]

Another proposal is to say that two noncognitive attitudes conflict when they cannot both be satisfied.[16] To evaluate whether this condition provides a successful account of disagreement, we need to consider what the relevant attitudes are more exactly (e.g., preferences, approval/disapproval, exhortations) and what would be required for them to be satisfied or frustrated.[17]

While there is much more that might be said in relation to these proposals, I will now move on to consider approaches to solving the problem of lost disagreement that appeal to the broader conversational setting in which moral utterances are made. Such accounts of disagreement appear well equipped to meet the challenge because they focus on the assumptions among interlocutors about their communicative exchange. These accounts might therefore have the resources to explain what the common topic is between interlocutors who assert compatible propositions but nevertheless intuitively disagree—where this is explained by appeal to mutual assumptions about the communicative

11 Harman, "Moral Relativism Defended"; and Finlay, "The Pragmatics of Normative Disagreement."

12 I talk about norms here, but one may talk about standards or ends instead.

13 Stevenson, *Ethics and Language*; Huvenes, "Disagreement Without Error"; and Björnsson and Finlay, "Metaethical Contextualism Defended."

14 Marques, "Doxastic Disagreement," 128.

15 Marques, "Disagreeing in Context," 6.

16 Stevenson, *Facts and Values*, 3; Björnsson and Finlay, "Metaethical Contextualism Defended," 27–28; and Ridge, *Impassioned Belief*.

17 For discussion and critique of this approach, see Marques, "Disagreeing in Context," 6–7; and Dreier, "Truth and Disagreement in Impassioned Belief," 457–58.

exchange instead of trying to find a clash in the semantic or pragmatic contents conveyed by the sentences that the interlocutors assert.[18] Since contextualists are committed to the claim that the semantic contents of sentences such as 1 and 2 are in fact not in tension, this approach seems like a fruitful way to explain why the speakers of these sentences nonetheless disagree.

In the next section, I discuss existing accounts of disagreement that follow this line of thinking and argue that they face serious problems. In section 3, I present a novel account and argue that it avoids the problems of the previous views and provides a forceful way for contextualists to make sense of different forms of disagreement.

## 2. OBJECTIONS TO PREVIOUS ACCOUNTS

The first account to be discussed is by David Plunkett and Timothy Sundell, according to which some disagreements can be understood as *metalinguistic negotiations* concerning how a term ought to be used.[19] For instance, when two speakers disagree about whether Jim is tall, they might disagree in virtue of having different opinions about how the term 'tall' is or ought to be used in the context. Similarly, speakers who use a moral term with different extensions might disagree because they are engaged in a metalinguistic negotiation about how a moral term ought to be used in the context. Hence, the disagreement is explained not by the semantic or pragmatic contents expressed by the speakers' utterances but rather by appealing to their metalinguistic commitments regarding what they are (or should be) talking about when they use the term 'wrong'.

The second kind of accounts that I discuss here are ones that adopt Robert Stalnaker's framework of communication. According to this framework, assertions aim to update the *common ground*, which consists of a set of propositions accepted as mutual belief among a group of interlocutors.[20] One such account of disagreement has been defended by Justin Khoo and Joshua Knobe.[21] According to them, speakers morally disagree when they propose incompatible

18 A problem for any account of disagreement that focuses on the broader setting of the communication is that people who have never engaged in conversation may disagree, which can be taken as evidence that disagreement is primarily about conflicting mental states. One response is to simply say that the accounts are restricted to disagreement in discourse. Another response is to develop a counterfactual account of disagreement, saying roughly that two speakers *would* disagree *if* they engaged in conversation with one another.

19 Marques, "Metalinguistic Negotiation and Speaker Error," 150; Plunkett and Sundell, "Disagreement and the Semantics of Normative and Evaluative Terms."

20 Stalnaker, "Common Ground," 704.

21 Khoo and Knobe, "Moral Disagreement and Moral Semantics."

norms to be added to the common ground (against which the moral sentences they utter will be assessed for truth). A similar account is defended by Alejandro Pérez Carballo and Paolo Santorio.[22] Their view is primarily presented as an account of how some forms of expressivism can make sense of communication as a kind of information exchange. Still, the view sheds light on how contextualists (and perhaps expressivists) may give a metalinguistic account of disagreement. According to Pérez Carballo and Santorio, speakers can disagree even if they adopt different norms, because when they engage in moral discussion, they presuppose it to be common ground that there is a norm that they ought to converge on.

All three of these accounts face two main problems. The first is a circularity worry. For the metalinguistic negotiation account, the issue is that one normative disagreement is reduced to another normative disagreement: the disagreement about what is *wrong* is reduced to a disagreement about how 'wrong' *ought* to be used. But we get the same disagreement problem at the metalinguistic level too. If the speakers accept different linguistic norms, then their disagreement about how the moral terms ought to be used is also lost.[23]

The Stalnakerian accounts also face this problem, because they similarly end up reducing moral disagreement to another normative disagreement. To see why, consider how Khoo and Knobe argue that the disagreement is about which norm *ought to be* added to the common ground, where to propose adding a norm to the common ground is a matter of *affirming* that norm, through "putting these norms forward as guides for living," communicating about some act that "we *should* not perform it, that we *should* feel guilt if we do, that we *should* encourage others to avoid doing it, and so on."[24] Similarly, Pérez Carballo and Santorio argue that even when interlocutors actually subscribe to different moral norms, they presuppose that there is a norm or standard "on which the participants' attitudes *ought to* converge."[25] But in much the same way as with the metalinguistic negotiation account, we now need an account of this further normative disagreement.

In response, proponents of these views might argue that there are two different *oughts* at play here—and their disagreement account is supposed to explain one but not the other. For instance, a proponent of the metalinguistic account might argue that moral disagreement is reduced to a disagreement about how

22 Pérez Carballo and Santorio, "Communication for Expressivists."

23 Eklund presents a version of this circularity problem in relation to the problems mentioned above in note 7 above (*Choosing Normative Concepts* in the chapter "Alternative Normative Concepts").

24 Khoo and Knobe, "Moral Disagreement and Moral Semantics," 131, 127 (emphasis added).

25 Pérez Carballo and Santorio, "Communication for Expressivists," 608.

'wrong' ought to be used, where this 'ought' is linguistic, not moral. While it is true that the disagreement is reduced to a linguistic 'ought' that is not moral, the question still arises how to account for disagreements in which this linguistic 'ought' is invoked. For contextualists about the linguistic 'ought', the same problem will arise again. There will be contexts where the meaning of the linguistic 'ought'—as it is used by different speakers—varies.

The second main problem is that the views misplace the topic of the disagreement. This worry is put forward by Cappelen, who argues that the metalinguistic negotiation view misrepresents what topic the speakers intuitively disagree about. The issue is that first-order moral disagreements are simply not intuitively about negotiating the meanings of words.[26] A related problem is raised by Marques, who argues that metalinguistic negotiation accounts of evaluative disagreement do not provide the right kind of disagreement. She argues that there are merely procedural linguistic reasons at stake rather than genuine evaluative reasons or norms.[27] In a nutshell, the problem is that the accounts provide us with the wrong kind of disagreement.

This problem can also be pressed for the Stalnakerian accounts. The first-order moral disagreement between conservatives and progressives is reduced to a disagreement about which norm to accept into the common ground.[28] But intuitively, the disagreement between the speakers of 1 and 2 is not about that. A conservative and a progressive may disagree about which is the correct moral norm, but they *also* disagree about whether abortion is wrong. These are two different disagreements. Hence, the Stalnakerian accounts also misconstrue the topic of disagreement as one about fundamental norms rather than about the moral status of abortion.

In addition to these two broader problems that affect all three views, there are also other more minor worries for the accounts. A worry about the metalinguistic negotiation account is that it seems to provide the same story for moral disagreement and merely verbal disagreement (such as the disagreement about tallness). Both types of disagreement are metalinguistic negotiations. But as I have argued in section 1, these are intuitively different. A similar criticism is raised by Marques in her discussion of Sundell's view of aesthetic and taste predicates as gradable adjectives.[29]

26 Cappelen, *Fixing Language*, 175.

27 Marques, "What Metalinguistic Negotiations Can't Do," 46.

28 One might argue that this view ultimately amounts to a metalinguistic negotiation account since the norm that is under negotiation is what determines the semantic content of the moral term.

29 Marques, "What Metalinguistic Negotiations Can't Do"; and Sundell, "The Tasty, the Bold, and the Beautiful."

A worry for the Stalnakerian proposals is that they rely on the thought that the common ground includes norms. But it is not obvious how to fit norms into Stalnaker's notion of common ground, which *by definition* consists of propositions accepted as mutual beliefs.[30] A similar worry has been raised by Rae Langton, as well as Marques and Manuel García-Carpintero, concerning the view that derogatory language adds practical contents (such as emotions or norms) to the common ground.[31] Marques also raises this problem in her criticism of the views by Khoo and Knobe and Pérez Carballo and Santorio. She argues that these views do not successfully provide "an account of shared acceptances of norms."[32] While this problem may not undermine these views, it demonstrates that there is much more to be said about the occurrence of norms in the common ground before such norms can play a crucial role in a satisfactory account of disagreement.

I will now present a new account of disagreement for contextualism and then argue that it solves the problems that have been raised in this section.

## 3. Moral Disagreement and Dynamic Pragmatics

According to dynamic pragmatics, there is a shared body of assumptions among a group of interlocutors that adjusts depending on the utterances that the interlocutors make. The most well-versed version of this idea—discussed in section 2 above—is Stalnaker's view that interlocutors make pragmatic presuppositions about what is common ground.[33] More recent developments in dynamic pragmatics provide a broader notion of a *discourse context*, which includes not only the common ground but also a *question set* and a *to-do list function*.[34]

30 For a positive proposal of how to solve this problem, see Björkholm, "Norms of Behavior and Emotions in the Discourse Structure."

31 Langton, "Beyond Belief," 85; and Marques and García-Carpintero, "Really Expressive Presuppositions and How to Block Them," 141. To be clear, Marques and García-Carpintero *defend* a view according to which derogatory utterances add reactive attitudes to the discourse context, but in defending this view, they criticize the option of saying this while maintaining the traditional Stalnakerian picture where the discourse context consists only of propositions accepted as common ground. They argue that to explain the conversational effects of derogatory claims, we need to accept a broader picture of the discourse structure that "include[s] at least a Stalnakerian common ground (the propositions that are accepted as true), QUD, and a set of plans" (141).

32 Marques, "Illocutionary Force and Attitude Mode in Normative Disputes," 460.

33 Stalnaker, "Common Ground," 704.

34 Portner, "The Semantics of Imperatives Within a Theory of Clause Types" and "Imperatives and Modals"; and Roberts, "Context in Dynamic Interpretation" and "Information

The to-do list function assigns properties onto an individual interlocutor's to-do list that represent acts or act-types that the interlocutor is committed to act in accordance with.[35] Note that to-do lists represent something other than mutual belief *about* what is permissible or required. The common ground may include the proposition that *it is required that* S *sits down*; and interlocutors might accept this proposition as part of the common ground, but *S* has not thereby *committed* to sitting down. It is when the act of sitting down is added to an interlocutor's to-do list that she publicly commits to performing the act (at least insofar as the other interlocutors are concerned). Put differently, a to-do list represents the acts an interlocutor is committed to perform, whereas the common ground can represent the beliefs they have about what the interlocutor is permitted or required to perform.[36]

The question set represents the *questions under discussion* (QUDs) of a conversation. By accepting a QUD into the question set, the interlocutors mutually accept that they aim to resolve it. The semantic contents of questions can be represented as the set of propositions that provide possible answers to it.[37] These are the *set of alternatives*. A QUD is resolved when the interlocutors successfully add one (or more) of the propositions among the set of alternatives to the common ground.

A *constituent* question such as "Who won Eurovision?" is represented by the set of (relevant) *polar* questions that constitute subquestions to it. For instance, the set of alternatives for "Who won Eurovision?" include the subquestions "Did Ukraine win Eurovision?", "Did Italy win Eurovision?", "Did Finland win Eurovision?", etc. The set of alternatives for such polar questions is represented by the propositions that provide *yes* and *no* answers to them. For instance, the set of alternatives for "Did Ukraine win Eurovision?" is "Ukraine won Eurovision" and "Ukraine did not win Eurovision."

In some cases, a constituent question is *partially* answered by answering one of the polar questions among the set of alternatives. For instance, a constituent question such as "Who competed in Eurovision?" is partially answered by answering the polar question "Did Ukraine compete in Eurovision?"—which is one of the polar questions among the set of alternatives. But resolving this polar

Structure in Discourse." Instead of to-do lists, one might use the notion of plan sets. See Han, "The Syntax and Semantics of Imperatives and Related Constructions."

35 Another option is to follow Ninan and represent a to-do list as a list of propositions (e.g., "*S* opens the door") that the interlocutor is committed to making true ("Two Puzzles About Deontic Necessity").

36 As Portner puts it, "at some point we have to form a commitment to act," which is the aspect of conversation that the to-do list models ("Imperatives and Modals," 381).

37 Karttunen, "Syntax and Semantics of Questions."

question does not provide a complete answer (since there are more countries than Ukraine that competed). By contrast, "Ukraine won Eurovision" would be a complete answer to the QUD "Who won Eurovision?"

I will now show how this framework can be used to provide an account of disagreement for contextualists. Bear in mind that like all models, this framework makes simplifications. It should be understood as a rational reconstruction of what happens in conversations and how the mutually accepted body of information is structured. The primary aim of the account presented here is to uncover the way that this formal framework can be used to represent what goes on in different cases of disagreement and to argue that it provides a way to understand what the common topic is, even when each speaker uses an expression with a different extension.

I will start by giving an account of disagreement that explains the different kinds of disagreement presented in section 1, namely, exclusionary disagreement, merely verbal disagreement, and disagreement without coextension—where moral disagreements between speakers who embrace different moral norms belong to the last category. I will argue that the QUD framework can be used to account for the differences between these kinds of disagreement. In section 3.2, I will then discuss how this account fares better regarding the objections presented in section 2 above.

### *3.1. The QUD Account of Disagreement*

The perhaps most straightforward way of thinking about a disagreement between two people who make seemingly conflicting assertions is that they are asserting incompatible propositions or hold incompatible beliefs. However, as Charles Stevenson famously points out, we can also conceive of a second type of disagreement—namely, disagreement in attitude—and, as briefly discussed in section 1, this thought has been developed in more detail in more recent debate. The proposal presented here can be understood as developing a third way of understanding disagreement—namely, as being over *questions*.

In a nutshell, I develop an account of what it is for the interlocutors who assert 1 and 2 to disagree about *whether abortion is wrong*, where this is explained as a disagreement over a mutually accepted question. In contrast to the approach of accounting for disagreement in belief or attitude, this questions-based approach does not aim to give an account of how the contents that are semantically or pragmatically expressed by two utterances are in conflict. Rather, the aim is to account for the way that two speakers who assert compatible propositions can disagree because they assert these propositions as answers to one and the same question that they aim to resolve. They are thus not speaking past each other.

The account crucially relies on the thought that interlocutors can accept QUDs that are *opaque*, which allows them to accept a common question even though they accept different moral norms. Before going into detail about what opaque questions are, I describe Stalnaker's related notion of *defective* contexts: a context is defective, according to Stalnaker, when the interlocutors' presuppositions about what propositions are in the common ground do not align, and it is nondefective when the interlocutors presuppose the same things.[38] In defective contexts, the interlocutors do not presuppose the same propositions in the common ground. According to Stalnaker, defective contexts might sometimes hinder efficient communication, but often, defects have little or no effect on communicative exchange.

In much the same way as the discourse context can be defective due to mistakes about what propositions are common ground, it can also be defective when interlocutors accept different QUDs. Interlocutors can make mistakes about whether a proposition is common ground, and they can make mistakes about whether a question is in the question set. Hence, the question set can be defective.

The notion of an opaque QUD captures something different from defective question sets. When interlocutors accept a nonopaque QUD, they have a shared idea of the meanings of their terms, whereas in opaque contexts, they do not. The difference between opaque and nonopaque QUDs is represented in the discourse context by means of the set of alternative answers. An opaque QUD has a wider set of propositions under each alternative answer than a nonopaque QUD has. To illustrate, consider first how a nonopaque polar QUD includes only one proposition under *Yes* and one under *No*. This can be represented as follows:

"Did Ukraine win Eurovision?"

| *Yes* | *No* |
|---|---|
| Ukraine won Eurovision | Ukraine did not win Eurovision |

FIGURE 1

In conversations between interlocutors who share the moral norms or standards that determine the meaning of their moral expressions, the interlocutors can accept a QUD that is nonopaque in much the same way as the QUD in figure 1 above. For instance, if two speakers who share moral norms disagree about whether abortion is wrong, they can have the following QUD (where *F* is the wrong-making property according to their shared norm):

38 Stalnaker, "Common Ground," 717.

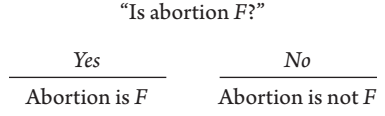


FIGURE 2

By contrast, in cases where the interlocutors do not subscribe to the same moral norm that determines the content of their moral expressions, the interlocturos do not have a clear *shared* idea about what property 'wrong' or 'right' refers to. However, they nevertheless engage in conversation with one another. In such cases, they accept an opaque QUD in which the expression 'wrong' must be understood as a placeholder. While the set of alternatives to a nonopaque QUD includes one proposition under *Yes* and one under *No*, it is characteristic of opaque QUDs that the set of alternatives includes a wider set of propositions than QUDs normally do when they are nonopaque. This can be represented in the following figure:

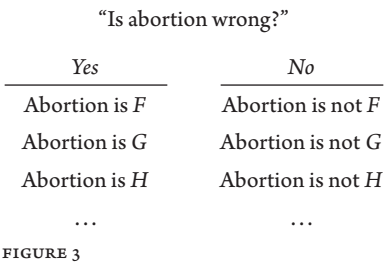


FIGURE 3

To clarify, when the moral question enters the discourse structure, it is represented as an opaque QUD in the way represented in figure 3. But it will still be true that for each speaker, if they were to utter the question "Is abortion wrong?" the semantic content of this sentence will be context sensitive in much the same way as when they make assertive moral utterances. But since conversations are understood as cooperative and rational goal-oriented endeavors according to the Stalnakerian picture, the content of the question cannot have this specific content, as it occurs as a mutually accepted QUD. The speakers take themselves to have a meaningful conversation about the wrongness of abortion even if they come to realize that they differ in their fundamental moral commitments. This must be reflected in their shared discourse structure in a way that respects them as rational and cooperative speakers. The QUD is therefore opaque as it occurs in the shared question set.

To illustrate, imagine again a conservative and a progressive who engage in a conversation about whether abortion is wrong. In this case, the progressive

is aware that the conservative does not share her deeper moral values, and vice versa. But they are nevertheless engaged in a conversation about whether abortion is wrong. To make sense of this, we can utilize the notion that a QUD can be opaque. It would be uncharitable to ascribe to the progressive the belief that they both accept a common inquiry of, for instance, trying to find out whether abortion respects a woman's right to autonomy over her own body, since the progressive knows that this is not reasonably what the conservative has in mind.[39]

With the distinction between opaque QUDs and defective contexts in place, it is now possible to explain how moral disagreement over opaque QUDs differs from exclusionary disagreement and merely verbal disagreement. In merely verbal disagreements, the question set is defective. The interlocutors' presuppositions about what QUDs are in the discourse context do not align. In such cases, they assume they have a shared (nonopaque) QUD but are mistaken about which QUD is mutually accepted. For instance, suppose that an interlocutor raises a QUD pertaining to whether Jim is tall. Since 'tall' is vague and reference-class relative, we can have discourses in which the speakers are talking past each other since they assume different contrast-classes for 'tall'. One interlocutor might think that the QUD concerns whether Jim is tall *for a twelve-year-old boy*, and the other assumes that the QUD concerns whether Jim is tall *for a grown-up man*. In this case, the speakers take themselves to share a QUD, but they are mistaken. They presuppose different (nonopaque) QUDs.

Hence, in merely verbal disagreements such as the one over whether Jim is tall, the interlocutors have a defective context in which one speaker accepts a QUD illustrated in figure 4, and the other speaker accepts the QUD represented in figure 5.

"Is Jim tall (for a twelve-year-old boy)?"

| *Yes* | *No* |
|---|---|
| Jim is tall (for a twelve-year-old boy) | Jim is not tall (for a twelve-year-old boy) |

FIGURE 4

39 It seems reasonable to think that there is at least some discrepancy between the deeper moral values of most interlocutors, even if not all cases are as extreme as the progressive and conservative case. We can think of cases in which speakers accept a nonopaque moral QUD as quite rare, since speakers will often not be aware of the moral norms that are accepted by the other. It makes sense to think of most moral conversations in which speakers are unsure about their interlocutor's moral values to be ones in which they accept opaque QUDs.

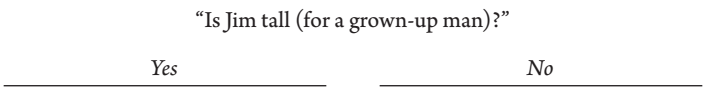


FIGURE 5

This is different from moral disagreement over opaque QUDs. In the disagreement about abortion between the progressive and the conservative, the *Yes* and *No* cells include more alternatives since the interlocutors mutually accept *the same* opaque QUD. When the speakers of 1 and 2 make their assertions 'abortion is wrong' and 'abortion is not wrong', they each propose different ways of answering this QUD—asserting propositions from the *Yes* and *No* cells, respectively. If they would accept either 1 or 2 into the common ground as an answer to the QUD, the question would be resolved and thus removed from the question set. But neither of them accepts what the other says into the common ground as an answer to their mutually accepted QUD. The opaque QUD remains in the question set unresolved, and their disagreement persists.

Note that the thought is not that an opaque moral QUD remains in the question set because the propositions asserted are merely *partial* answers to it. Rather, each proposition among the set of alternatives should be understood as a *complete* answer to the opaque QUD. Hence, the account presented here is not that the propositions in the *Yes* and *No* cells provide partial answers to the opaque QUD of whether abortion is wrong. Rather, the interlocutors disagree because they accept a common QUD and propose different propositions as complete answers to it, but neither of them accept that the proposition the other asserts resolves their mutually accepted QUD.

One might push back on this way of distinguishing between moral disagreement and merely verbal disagreement, arguing that if contextualism is correct and the speakers assert propositions that concern different things, they do not seem to have a common QUD after all. Instead, each of their respective utterance will be considered off topic from the perspective of the other, since they do not consider the norm accepted by the other to be relevant for determining the moral value of abortion.[40] For instance, the properties that are relevant according to the progressive's moral norm will be irrelevant to answering the QUD of whether abortion is wrong, according to the conservative. Hence, it turns out to be a merely verbal disagreement or a case in which they simply speak past each other.

40 Thanks to an anonymous referee for pointing me to this objection.

In response, it is important to highlight two things. First, it is because of the cooperative nature of conversation that it is reasonable to think that the QUD accepted by the conservative and the progressive is an opaque one. So we want to make sense of the way that the speakers engage in a common inquiry, and the notion of opaque QUDs is designed to make sense of this.

Second, since both propositions asserted are part of the set of alternatives to this opaque QUD, neither of them says something that is off topic. It is correct that neither of the speakers thinks that the proposition asserted by the other is *the* answer to the QUD, but each can still accept that the proposition asserted by the other is a possible answer among the set of alternatives.

To compare, suppose you engage in a conversation with someone about who won Eurovision, and you believe that Ukraine won Eurovision. Even though you think that this is the answer to the QUD, you can still accept that there are other propositions that constitute possible answers (and propositions that do not). For instance, you accept that "Italy won Eurovision" and "Finland won Eurovision" are other alternative answers (but "Costa Rica won Eurovision" is not). Similarly, the progressive can accept that the proposition that she asserts is the answer to the opaque moral QUD and yet accept that the proposition asserted by the conservative is another possible answer among the set of alternatives. Insofar as they accept an opaque QUD, neither of their assertions is off topic.

Hence, the account presented here can explain the difference between merely verbal disagreements and moral disagreement without coextension. Moreover, it also distinguishes disagreement without coextension from exclusionary disagreement. In exclusionary disagreements, such as when the two speakers disagree about whether Ukraine won Eurovision, they accept the QUD from figure 1, to which there can be only one true answer. Similarly, in a moral disagreement where the interlocutors accept the same norm, the interlocutors can accept a QUD such that the *yes* and *no* cells include only the propositions "abortion is *F*" and "abortion is not *F*" (represented in figure 2). Since abortion cannot both be *F* and not be *F*, their disagreement is exclusionary. Hence, some moral disagreements may be of this kind.

In moral disagreement where the speakers do not share their fundamental moral norms, such as the disagreement between the progressive and the conservative, they will instead accept an opaque QUD. In such cases, the disagreement is not exclusionary because it will be true that the interlocutors *could* accept the proposition that the other asserts into the common ground without being inconsistent: they could accept both that abortion is *F* and that abortion is not *G* without inconsistency. But they still disagree because they do not accept what the other has said *as an answer* that resolves the opaque QUD of whether abortion is wrong. The contents of both 1 and 2 are members of the set

of alternatives to the mutually assumed QUD. The disagreement persists since they cannot coordinate on a proposition to add to the common ground that would resolve and remove the QUD from the question set. The QUD remains unresolved by interlocutors, even though they offer answers to it that are members of the set of alternatives.

One might object to this view of disagreement by arguing that it would be uncharitable to interpret speakers who endorse *moral realism* of accepting opaque QUDs (where moral realism is the view that the meaning of 'wrong' is context invariant and that sentences in which it occurs can be objectively true).[41] Interlocutors who subscribe to different moral norms but accept moral realism might be more likely to think that they accept a nonopaque QUD of whether abortion has *the* (context-invariant) property of being wrong—to which the set of alternatives are simply "abortion is wrong" and "abortion is not wrong".[42]

One way to make sense of this example within the framework is to say that while these interlocutors think that they have a nonopaque QUD in common, the meaning of their utterances require of the discourse structure that their QUD is opaque. To clarify, if we presuppose that contextualism is correct, it will still be the case that two interlocutors who believe moral realism to be true while endorsing different moral norms will pick out different properties by their respective utterances of 'wrong'. Although they think that they are referring to the same property, they refer to different properties insofar as contextualism is true about the meaning of their claims. Since they take themselves to have a common QUD, this QUD must be represented in the discourse structure as an opaque one since the contents of their respective utterances would otherwise not be answers to the same QUD. It is therefore charitable to interpret their QUD as an opaque one to accommodate the way that they take themselves to have a common inquiry.[43]

The QUD account of moral disagreement presented here can account for the empirical data supporting that moral disagreements between people who share moral norms are exclusionary, whereas disagreements between people with

41 We can think not only of speakers who explicitly accept the philosophical position of moral realism but also of speakers who more implicitly accept the view that moral facts are objectively true.

42 Thanks to an anonymous referee for pressing this concern.

43 This example raises interesting and difficult questions about how interlocutors' beliefs about what they are referring to affects the contents of the discourse structure. In general, do speakers update the common ground with the proposition they think they are asserting, or the proposition that they are actually asserting? If we assume an externalist semantics, these two are not necessarily the same, since speakers can be mistaken about what the contents of their utterances are (if the content is determined by factors other than the speakers' intentions). While I find this issue interesting, I think it goes beyond the topic of this paper.

different moral norms are not.[44] In contexts where speakers share moral norms, they have nonopaque QUDs to which there is one objectively true answer. Only one of the interlocutors can be correct. But when they share an opaque QUD, there is not one objectively correct answer, even though they strive to coordinate on one answer.

One might object to this, wondering why interlocutors strive to coordinate on one answer over opaque moral QUDs if the propositions they assert can both be true. Again, if we want to construe the interlocutors as rational, cooperative and goal-oriented, the question arises why they are inclined in this case to accept an opaque QUD at all and to coordinate on an answer to it.

One way one might try to explain this is to say that opaque moral QUDs have a certain *normative role*, and this puts pressure on coordinating on an answer to them. But what is this normative role? Matti Eklund characterizes the normative role of a moral expression as the way it is "fit to be used in practical deliberation about what to do" and how its "application has, so to speak, practical consequences in addition to merely theoretical ones."[45] While I agree with Eklund that there is intuitively something like a normative role, the characterization of what a normative role is needs to be expanded.

It is possible to make this idea more precise within the dynamic pragmatic framework. In broad strokes, the proposal is that moral QUDs have a *practical* role that resides in the way they are part of an (implicit) series of QUDs that originate from the overarching question of *what to do*. To unpack what this means, I will clarify both what an *overarching question* is and what it is to be *part of a series of QUDs*. I will explain these two things in turn.

First, in the standard Stalnakerian model, the broad overarching QUD of rational goal-oriented discourse is thought of as: *What is the world like?* This QUD is taken as a background presupposition for most (perhaps all) discourse. But as Paul Portner contends, "Conversation is also about planning and coordinating action."[46] Hence, we can combine the claims from Stalnaker and Portner by saying that there are two overarching QUDs of rational discourse: one concerning what the world is like and one concerning what to do. The role of the common ground and the to-do list function is to track answers to these two broad questions of discourse.

Second, the question set helps narrow in on more specific QUDs that are taken to be conducive to answering these broad questions, thus creating a series of interrelated QUDs. Craige Roberts maintains that when interlocutors have accepted

44 I am grateful to Isidora Stojanovic for suggesting that this is a benefit of the account.

45 Eklund, "Alternative Normative Concepts," 152–53.

46 Portner, "Imperatives and Modals," 381.

a QUD, they can either provide an answer to it or introduce a subquestion to it.[47] For instance, if the QUD is "Who won Eurovision?" an interlocutor might either say "Ukraine won Eurovision" (thereby asserting an answer) or "Did Ukraine win Eurovision?" (thereby introducing a subquestion). In addition, Roberts also acknowledges that interlocutors may add new QUDs that are conducive to answering a former one in virtue of assumptions of what else is in the discourse context.[48] I will call these *auxiliary* questions.[49] For instance, in response to the QUD of who won Eurovision, an interlocutor might ask "Did France compete this year?" which is considered relevant in virtue of the assumption that only countries that competed could have won. And thus, an answer to this auxiliary question has the potential to narrow down the set of possible alternative answers.

The notion of a practical or normative role can now be spelled out. The thought is that QUDs concerning what is wrong or right have a role in an implicit series of QUDs that originate in the broad practical QUD of what to do. Since figuring out what the world is like is crucial to be able to plan and coordinate actions, the question of what the world is like will always be an auxiliary question to the question of what to do (but not vice versa). QUDs that are taken to be auxiliary include more specific questions about what is right, wrong, or what one ought to do. It is thus part of the background assumptions of moral discourse that interlocutors want to settle a broad overarching QUD concerning what to do, and questions concerning what is right or wrong are auxiliary to answering it. Hence, the QUD "Is abortion wrong?" is not raised from nowhere—it is raised as a QUD that is assumed to be conducive to answering the broader overarching QUD about what to do. Therefore, the interlocutors aim to settle on one answer to this opaque moral QUD that is conducive to answering the overarching practical question of what to do.

To sum up, I have argued that moral disagreements can be about opaque moral QUDs. The interlocutors disagree because they both accept an opaque QUD pertaining to whether abortion is wrong, but neither speaker accepts the proposition asserted by the other into the common ground as an answer to their shared opaque QUD. The interlocutors are talking about the same thing since they are both aiming to resolve the same opaque QUD, and the contents of their assertions are members of the set of alternatives to that QUD. But neither of their asserted contents is accepted as an answer that removes the QUD from the question set. This account of disagreement allows contextualists to preserve their core commitment that when two speakers use moral expressions, the contents of

47 Roberts, "Information Structure in Discourse," 6–7.

48 Roberts, "Information Structure in Discourse," 7.

49 Björkholm, "The Duality of Moral Language," 135.

their claims vary across contexts of utterance and can therefore both be true. But at the mutually presupposed discourse level, the speakers accept a QUD to which the contents of both their assertions are potential answers—and the pressure to settle on one answer to this QUD resides in the way the question is part of an implicit series of QUDs that trace back to an overarching question of what to do.

### *3.2. Return to the Objections*

Before I conclude, let us briefly return to the objections raised in section 2 to see how the account presented here can avoid the same pitfalls. I discussed two main objections, as well as two more specific worries for each of the views. I start here by briefly discussing the specific worries. First, I raised the worry that the Stalnakerian accounts rely on norms being part of the common ground, even though the common ground is by definition designed to model common belief. This exact objection is avoided by the account presented here since I do not explain disagreement by appealing to norms in the common ground and have therefore not incurred a burden to explain how norms are represented in the discourse context. Still, my account requires an explanation for how there can be *questions* in the discourse context—and subsequently does so by appealing to and further expanding the influential QUD framework.

Second, I argued that Plunkett and Sundell's account does not accommodate the difference between disagreement without coextension and merely verbal disagreements. The account presented here can distinguish moral disagreements from merely verbal disagreements. In the former case, interlocutors know that they employ a moral expression that they may assign very different meanings to, but they still accept a QUD that is therefore opaque. In the latter case, the speakers think that they mutually accept a nonopaque QUD, but they are mistaken. Once they realize that their question set is defective, their disagreement will dissolve upon acknowledgement that they did not have the same QUD in mind. The account does therefore not overgeneralize to cases in which interlocutors use context-sensitive expressions, but their disagreement is intuitively merely verbal. Hence, the QUD account captures the difference between merely verbal disagreements and disagreement without coextension by appealing to the difference between defective and opaque question sets.

I now turn to the two more serious problems discussed in section 2. Recall that the problem of misplacement is the objection that the first-order moral disagreement about whether abortion is wrong is reduced to the wrong kind of disagreement. As Cappelen argues, Plunkett and Sundell's account locates disagreement in the wrong place—that is, as disagreement over the meanings of words rather than as first-order moral disagreement. I have argued that the

Stalnakerian accounts also face a similar worry since they reduce the first-order disagreement to a second-order disagreement about what norms to accept.

This objection is avoided by the QUD account. The disagreement is neither reduced to a disagreement over the meanings of words nor to a deeper disagreement about what moral norms to accept. According to the account presented here, the speakers of 1 and 2 disagree about *whether abortion is wrong* even if they accept different norms—because they accept an opaque QUD. Their disagreement is about the question of whether abortion is wrong, and they fail to resolve their disagreement because neither accepts the proposition asserted by the other as an answer that removes the QUD from the question set. This is an intuitive way of thinking about disagreement as being over first-order moral questions.

The circularity problem is that the views explain one normative disagreement by reducing it to another normative disagreement. The problem for the metalinguistic negotiation account is that it reduces moral disagreement to a normative disagreement about how a normative term *ought to be used*, and the Stalnakerian views reduce it to a disagreement about which norm or standard the interlocutor's *ought* to converge on. The QUD account avoids this problem too since it does not claim that moral disagreement is reduced to a more fundamental normative disagreement. Rather, I have argued that in moral disagreements without coextension, the speakers accept an opaque QUD and disagree on which proposition among the set of alternatives resolves this question.

One might object that although the QUD account does not face exactly the same circularity worry as the other accounts, there is still some normativity in the background since I argued that we might explain the pressure for coordinating on one answer to the opaque QUD by referring to its role in a series of QUDs conducive to answering the broad overarching question of what to do. However, it is important to note that the circularity objection is not that there cannot be any normativity involved in the background of a normative disagreement. Rather, the problem is that the disagreement between the speakers of 1 and 2 should not be reduced to just another normative disagreement. Even though the question of what to do is the origin of the series of QUDs that precede the opaque QUD of whether abortion is wrong, the disagreement between 1 and 2 is not reduced to a disagreement about this overarching question.

Moreover, it can be contended that the question of what to do need not be reduced to a question of what one *ought* to do.[50] While Allan Gibbard treats the question of what to do as identical to the question of what one ought to

50 One might object that this distinction can also be made by proponents of the other metalinguistic view, and they might therefore also avoid the circularity worry. Although it might be true that they can, I think it is up to the proponents of these views to show how this distinction can be utilized to aid their respective accounts.

do, both Olle Risberg and Justin Clarke-Doane argue that they are distinct.[51] Risberg follows Gibbard's approach of treating the question of what to do as one being answered by forming *intentions* of how to act, but in contrast to Gibbard, he maintains that questions of what one ought to do (prudentially, morally, all things considered, etc.) are all questions that can be given true answers. Clarke-Doane argues that the question of what to do is answered by forming a noncognitive attitude. While I agree with Risberg and Clarke-Doane that answering the question of what to do cannot be fully answered by true propositions, my view is not that it is instead answered by forming intentions or noncognitive attitudes. Rather, as I have argued, this fundamentally practical question is answered by adding actions to to-do lists (which may in turn commit interlocutors to have certain intentions).

In short, the QUD account succumbs neither to the circularity problem nor to the misplacement problem, since it does not reduce moral disagreement to another kind of normative disagreement.

## 4. CONCLUSION

I have argued that interlocutors can morally disagree even when they use moral expressions with different extensions. They mutually accept an opaque QUD, and both assert propositions that are members of the set of alternatives to this QUD. But while both speakers make their assertions as intended complete answers to the QUD, neither accepts what the other says as an answer that removes it from the question set. This account allows contextualists to preserve their core semantic claim that when speakers use moral expressions, the semantic contents of their claims vary across contexts of utterance. But at the mutually presupposed discourse level, the speakers presuppose a common opaque QUD, to which the contents of both their assertions are among the set of alternatives. Hence, there is one sense in which they are talking about different things and another sense in which they are talking about the same thing: they refer to different properties but accept a mutual question. In this sense, they are talking about the same topic even though their moral expressions are not coextensive.[52]

*Institute for Futures Studies*
*Stockholm University*
*stina.bjorkholm@iffs.se*

51 Gibbard, *Thinking How to Live*; Risberg, "Ethics and the Question of What to Do"; and Clarke-Doane, *Morality and Mathematics*.

52 I am especially grateful to Anandi Hattiangadi, Anders Schoubye, Gunnar Björnsson and Teemu Toppinen for helpful discussions and comments on early drafts of this paper. I

## REFERENCES

Beebe, James R. "Moral Relativism in Context." *Noûs* 44, no. 4 (2010): 691–724.

Björkholm, Stina. "The Duality of Moral Language: On Hybrid Theories in Metaethics." Doctoral diss., Stockholm University, 2022. http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-203526.

———. "Norms of Behavior and Emotions in the Discourse Structure." *Journal of Social Ontology*, forthcoming.

Björnsson, Gunnar. "The Significance of Ethical Disagreement for Theories of Ethical Thought and Talk." In *Routledge Handbook of Metaethics*, edited by Tristram McPherson and David Plunkett. Routledge, 2017.

Björnsson, Gunnar, and Stephen Finlay. "Metaethical Contextualism Defended." *Ethics* 121, no. 1 (2010): 7–36.

Brendel, Elke. "Contextualism, Relativism, and the Problem of Lost Disagreement." In *Routledge Handbook of Epistemic Contextualism*, edited by Jonathan Jenkins Ichikawa. Routledge, 2017.

Cappelen, Herman. *Fixing Language: An Essay on Conceptual Engineering*. Oxford University Press, 2018.

Cappelen, Herman, David Plunkett, and Alexis Burgess. *Conceptual Engineering and Conceptual Ethics*. Oxford University Press, 2019.

Clarke-Doane, Justin. *Morality and Mathematics*. Oxford University Press, 2020.

Cohnitz, Daniel, and Teresa Marques. "Disagreements." *Erkenntnis* 79, no. 1 (2014): 1–10.

Dreier, Jamie. "Truth and Disagreement in Impassioned Belief." *Analysis* 75, no. 3 (2015): 450–59.

Eklund, Matti. "Alternative Normative Concepts." *Analytic Philosophy* 53, no. 2 (2012): 139–57.

———. *Choosing Normative Concepts*. Oxford University Press, 2017.

Finlay, Stephen. "The Pragmatics of Normative Disagreement." In *Having It Both Ways: Hybrid Theories and Modern Metaethics*, edited by Guy Fletcher and Michael Ridge. Oxford University Press, 2014.

Gibbard, Allan. *Thinking How to Live*. Harvard University Press, 2003.

Goodwin, Geoffrey P., and John M. Darley. "The Perceived Objectivity of Ethical Beliefs: Psychological Findings and Implications for Public Policy."

have presented versions of this paper at the Graz Evaluative Language Workshop (2023), the SIFA Midterm Conference on Truth in Evaluation in Alghero (2023), and the Swedish Congress of Philosophy in Lund (2022). I am grateful to the organizers and participants of these events for the helpful feedback I received. Finally, many thanks to the anonymous reviewers of this journal. The research was partly funded by the Swedish Research Council (grant no. 2019-03347).

*Review of Philosophy and Psychology* 1, no. 2 (2010): 161–88.
———. "The Psychology of Meta-Ethics: Exploring Objectivism." *Cognition* 106, no. 3 (2008): 1339–66.
———. "Why Are Some Moral Beliefs Perceived to Be More Objective than Others?" *Journal of Experimental Social Psychology* 48, no. 1 (2012): 250–56.
Han, Chung-Hye. "The Syntax and Semantics of Imperatives and Related Constructions." PhD diss., University of Pennsylvania, 1998.
Harman, Gilbert. "Moral Relativism Defended." *Philosophical Review* 84, no. 1 (1975): 3–22.
Huvenes, Torfinn Thomesen. "Disagreement Without Error." *Erkenntnis* 79, no. 1 (2014): 143–54.
———. "Varieties of Disagreement and Predicates of Taste." *Australasian Journal of Philosophy* 90, no. 1 (2012): 167–81.
Karttunen, Lauri. "Syntax and Semantics of Questions." *Linguistics and Philosophy* 1, no. 1 (1977): 3–44.
Khoo, Justin. "The Disagreement Challenge to Contextualism." In *Routledge Handbook of Epistemic Contextualism*, edited by Jonathan Jenkins Ichikawa. Routledge, 2017.
Khoo, Justin, and Joshua Knobe. "Moral Disagreement and Moral Semantics." *Noûs* 52, no. 1 (2018): 109–43.
Kölbel, Max. "Faultless Disagreement." *Proceedings of the Aristotelian Society* 104, no. 1 (2004): 53–73.
Langton, Rae. "Beyond Belief: Pragmatics in Hate Speech and Pornography." In *Speech and Harm: Controversies over Free Speech*, edited by Ishani Maitra and Mary Kate McGowan. Oxford University Press, 2012.
MacFarlane, John. "Relativism and Disagreement." *Philosophical Studies* 132, no. 1 (2007): 17–31.
Marques, Teresa. "Disagreeing in Context." *Frontiers in Psychology* 6 (2015): 1–12.
———. "Doxastic Disagreement." *Erkenntnis* 79, no. 1 (2014): 121–42.
———. "Illocutionary Force and Attitude Mode in Normative Disputes." *Metaphilosophy* 52, nos. 3–4 (2021): 449–65.
———. "What Metalinguistic Negotiations Can't Do." *Phenomenology and Mind* 12 (2017): 40–48.
Marques, Teresa, and Manuel García-Carpintero. "Really Expressive Presuppositions and How to Block Them." *Grazer Philosophische Studien* 97, no. 1 (2020): 138–58.
Ninan, Dilip. "Two Puzzles about Deontic Necessity." In *New Work on Modality*, edited by J. Gajewski, V. Hacquard, B. Nickel, and S. Yalcin. MIT Working Papers in Linguistics, 2005.

Pérez Carballo, Alejandro, and Paolo Santorio. "Communication for Expressivists." *Ethics* 126, no. 3 (2016): 607–35.

Plunkett, David, and Timothy Sundell. "Disagreement and the Semantics of Normative and Evaluative Terms." *Philosophers' Imprint* 13, no. 10 (2013): 1–37.

———. "Metalinguistic Negotiation and Speaker Error." *Inquiry* 64, nos. 1–2 (2021): 142–67.

Pölzler, Thomas. "Revisiting Folk Moral Realism." *Review of Philosophy and Psychology* 8, no. 2 (2017): 455–76.

Pölzler, Thomas, and Jennifer Cole Wright. "Anti-Realist Pluralism: A New Approach to Folk Metaethics." *Review of Philosophy and Psychology* 11, no. 1 (2020): 53–82.

———. "An Empirical Argument Against Moral Non-Cognitivism." *Inquiry* 66, no. 6 (2020): 1–29.

Portner, Paul. "Imperatives and Modals." *Natural Language Semantics* 15, no. 4 (2007): 351–83.

———. "The Semantics of Imperatives Within a Theory of Clause Types." In *Proceedings of SALT 14*, edited by Robert B. Young. Linguistic Society of America, 2004.

Ridge, Michael. *Impassioned Belief*. Oxford University Press, 2014.

Risberg, Olle. "Ethics and the Question of What to Do." *Journal of Ethics and Social Philosophy* 25, no. 2 (2023).

Roberts, Craige. "Context in Dynamic Interpretation." In *Handbook of Pragmatics*, edited by Laurence R. Horn and Gregory Ward. Blackwell, 2006.

———. "Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics." *Semantics and Pragmatics* 5 (2012): 1–69.

Sarkissian, Hagop. "Aspects of Folk Morality: Objectivism and Relativism." In *A Companion to Experimental Philosophy*, edited by Justin Sytsma and Wesley Buckwalter. Wiley Blackwell, 2016.

Stalnaker, Robert C. "Common Ground." *Linguistics and Philosophy* 25, nos. 5–6 (2002): 701–21.

Stevenson, Charles L. *Ethics and Language*. American Mathematical Society, 1945.

———. *Facts and Values*. Yale University Press, 1963.

Streumer, Bart. *Unbelievable Errors: An Error Theory About All Normative Judgments*. Oxford University Press, 2017.

Sundell, Tim. "The Tasty, the Bold, and the Beautiful." *Inquiry* 59, no. 6 (2016): 793–818.

Tersman, Folke. "Moral Disagreement." In *International Encyclopedia of Ethics*, edited by Hugh LaFollette. John Wiley and Sons, 2016.

———. "Moral Disagreement." In *Stanford Encyclopedia of Philosophy* (fall 2022). https://plato.stanford.edu/archives/fall2022/entries/disagreement-moral/.

Wright, Jennifer C., Piper T. Grandjean, and Cullen B. McWhite. "The Meta-Ethical Grounding of Our Moral Beliefs: Evidence for Meta-Ethical Pluralism." *Philosophical Psychology* 26, no. 3 (2013): 336–61.

Zeman, Dan. "Faultless Disagreement." In *Routledge Handbook of Philosophy of Relativism*, edited by Martin Kusch. Routledge, 2020.

*Journal of Ethics and Social Philosophy*
VOL. 30, NO. 2 · APRIL 2025
*https://doi.org/10.26556/jesp.v30i2.3970*
© 2025 Author

# ABUSE OF POWER

*Dale Dorsey*

THE ELECTION of Barack Obama to the presidency of the United States left Illinois with a vacant senate seat. Under Illinois law, then-Governor Rod Blagojevich was charged with appointing someone to fill the remainder of Obama's term. According to investigators, Blagojevich attempted to sell that seat to the highest bidder. His scheme was discovered, and he was duly impeached by the Illinois General Assembly. The first article of impeachment reads:

> Based on the totality of the evidence contained in the Record of the House Special Investigative Committee created under House Resolution 1650 . . . in his conduct while Governor of the State of Illinois, Rod R. Blagojevich has abused the power of his office in some or all of the following ways: (1) The Governor's plot to obtain a personal benefit in exchange for his appointment to fill the vacant seat in the United States Senate.[1]

These impeachment proceedings begin with the accusation of *abuse of power*. This is far from rare in such proceedings. The impeachments (or near impeachments) of Richard Nixon, Donald Trump, and a number of other officials specifically cite abuse of power as among the reasons for impeachment and removal from office.

But reference to abuse of power goes far beyond the quasi-legal proceedings cited here. We accuse doctors, attorneys, police officers, and so on of abusing their power. But what is the abuse of power? How does it differ from other forms of power exercise? How does it differ from a simple commission of immoral, unjust, or otherwise impermissible behavior? What are the normative dimensions of power abuse? Must it be morally illegitimate or unjust? Illegitimate in some other way? Is abuse of power just a chimera? Note that not

1 Illinois General Assembly, 95th General Assembly (2009–11), House Resolution 1671 ("Impeachment of Governor"), https://www.ilga.gov/legislation/BillStatus.asp?DocNum=1671&GAID=9&DocTypeID=HR&LegId=40049&SessionID=51&GA=95.

all public officials who are, e.g., impeached, are charged with abuse of power.[2] This raises the possibility that abuse of power is just what we call something when we do not like what power holders have done but cannot figure out what else to call it.

In this paper, I aim to accomplish two things. First, I offer an analysis of the abuse of power that roughly defines power abuse as the exercise of power by one individual over another (or others) outside of the *raison d'être* of the power relation that they occupy. Second, I argue that, conceived in this way, individuals have (not necessarily decisive) practical reasons to avoid abusing power. Generally, then, the use to which the notion of power abuse is put in public life is not a normative chimera. One abuses power when one steps beyond the point of one's power relation over others, the point of which power holders have practical reason to observe.

The plan of the paper is as follows. After discussing a set of preliminary matters in section 1, I begin my analysis of the abuse of power in section 2 with a discussion of the notion of power at issue. In section 3, I discuss various accounts of how this power might be abused and argue for a relationship-centric account: i.e., what constitutes an abuse of power must be determined by the point of the power relation itself. After discussing an important objection to this account, I then consider the normative significance of power abuse in sections 4 and 5. I argue that individuals face reason not to abuse power irrespective of the justice, morality, or overall justifiability of such abuse.

## 1. PRELIMINARIES

Before I begin my examination, I should say a bit more to triangulate the concept I am interested in discussing. Arguably, any wrong behavior could be trivially dubbed an abuse of power, insofar as doing anything at all implies a kind of power (i.e., the power to do that thing) and insofar as wrong behavior is a *mis*use of that kind of power. But this is not typically what is meant by the term 'abuse of power'. When a president is impeached for abuse of power, when an athletic director is fired by a university for abuse of power, and so on, we do not mean that he or she acted in any old way that could be classified as rude, imprudent, or immoral. Rather, we seem to be referring to a particular kind

2 Donald Trump, for instance, in his second impeachment was charged with "incitement of insurrection" but not with abuse of power. See United States House of Representatives, 117th Congress (2021–22), House Resolution 24 ("Impeaching Donald John Trump, President of the United States, for High Crimes and Misdemeanors"), https://www.congress.gov/bill/117th-congress/house-resolution/24/committees.

of abuse—abuse that is tied specifically to the power of the position he or she holds over others and the expectations of behavior given that position.

Often when we talk about abuse of power, we mean to focus our attention on holders of some public office or others that maintain some sort of "official" status: governors who sell senate seats to the highest bidder, judges who find defendants guilty for issuing public criticism, presidents who blackmail dependent allies into dishing dirt on political opponents, prime ministers who prorogue parliament for the sake of avoiding parliamentary scrutiny of controversial policies, and so forth.[3] But there is no clear reason why we need to hold that abuse of power *must*, at least as a conceptual matter, be tied specifically to "officials" in this way. Employers, even if they are not public officials, can clearly abuse their power over employees.[4] A physician can abuse power over a patient. A professor can abuse power over students.

Relatedly, I wish to distinguish abuse of power from *corruption*. Though power abuse and corruption can clearly overlap and while certain accounts of corruption (see section 3.5 below) have some similarities with power abuse, these concepts are not plausibly synonymous or coextensive.[5] Of course, the proper analysis of corruption is contested, but intuitively speaking, there appears to be substantial daylight between abuse of power and corruption. A president who uses his position to seduce an impressionable intern is plausibly abusing power but is not, or at least need not be, corrupt. And while there is considerable dispute concerning the notion of corruption, many (though not all) hold that corruption must feature some sort of *moral* degradation, which may or may not be present (as I argue later) in cases of power abuse.[6]

3 See, e.g., the impeachment of District Court Judge James Peck (*Journal of the United Staters House of Representatives*, 21st Congress (1829–30), April 27, 1830); and article 1 of the first impeachment of President Donald Trump (United States House of Representatives, 116th Congress (2019–20), House Resolution 755 ("Impeaching Donald John Trump, President of the United States, for High Crimes and Misdemeanors"), https://www.congress.gov/bill/116th-congress/house-resolution/755). Note that there may be a more general *legalistic* sense of abuse of power that is relevant here. I am ignoring that for present purposes and simply inquiring whether there is a normatively significant notion of power abuse that the legalistic sense may or may not track. Thanks to an anonymous reviewer.

4 See, for instance, Vredenburgh and Brender, "The Hierarchical Abuse of Power in Work Organizations."

5 Notwithstanding an important definition of corruption (Nye, "Corruption and Political Development") that identifies it as "abuse of power for personal gain." For independent critique of this suggestion, see Miller, "Corruption."

6 Miller identifies corruption as a "species of immorality" ("Corruption"). For a case of power abuse that does not entail moral degradation, see the case of the medical skeptic in section 3.3 below. Thanks to an anonymous reviewer for prompting this discussion.

In genuine instances of power abuse, we seem to attribute a special kind of failure that stems from the power that is possessed by the abusers. And it is this sense of abuse of power—the intuitive idea of which I hope will be straightforward—that I investigate here.

## 2. ABUSE OF POWER, PART ONE: POWER

To begin, abuse of power requires power to be abused. But what is *power* of this kind? Abuse of power as it is commonly understood seems to make reference to power in social and interpersonal forms. Let us take the second clause first. Typical examples of power abuse seem to indicate that to abuse power is to abuse a particular form of power that one maintains *with regard to or over some other person or persons*—i.e., power that exists in the context of interpersonal relations. Indeed, there is a straightforward rationale for this point: if use of power in some particular way has no effect on any other person, it is hard to see how it can rightly be said to be an *abuse*.

The relationality of power is clear from typical examples—employee/employer, president/dependent ally, governor/constituents, physician/patient, lawyer/client, and so on. The first party has power *over* another or others. Now, for one person to have power *over* another calls out for analysis. For my purposes and for the sake of brevity, to say that a person has power over someone else is to say that she has power to influence that person or persons.[7] Of course, influence of this kind can take many forms. The clearest example of this kind of power is the power to influence action.[8] The power I have to compel behavior on the part of others is liable to abuse if, say, I compel a dependent ally to dish political dirt on my opponent or if I compel a student to wash my car on pain of a failing grade. But compelling power is not the only form of power that is relevant here. Consider, for instance, power of attorney. If my client is in a coma, I have no capacity to compel behavior. But I still have power to materially affect the aims, interests, and welfare of my client. Such power can clearly be abused—if, for instance, I used my client's money to pay off my gambling debts.[9] Indeed, while Blagojevich possessed *some* forms of power to compel his constituents, the exercise of power in selling a senate seat was not compelling power but

7 Nothing in this analysis is committed to the view, expressed by Brian Fay, that power of this kind is *necessarily* dyadic—i.e., a power holder versus the person over whom that power is held. See Fay, *Critical Social Science*, 120.

8 Dahl, "The Concept of Power."

9 Abuse of power of attorney is generally understood as the use of power of attorney against the best interests of the client. See, for instance, Stiegel, "Durable Power of Attorney Abuse."

rather, as we might say, *establishment power*: power to install an individual in a senate seat on his constituent's behalf.[10] "Influence," when it comes to power, should therefore be interpreted broadly; person *X* has power over person *Y* to the extent that *X* can influence the behavior, well-being, material interests, and so on, of *Y*. Note that these power relations need not be unidirectional: *X* might have power over *Y*, and *Y* over *X*, (as in, e.g., a joint fiduciary arrangement). But this sort of bidirectional power is surely subject to abuse.

Abuse of power thus makes reference to interpersonal power relations. But this power is also *social* in nature. In particular, abuse of power seems to require forms of power that are embedded within *social institutions*. For instance, Blagojevich abused his power over his constituents insofar as he was the governor of Illinois; employers abuse their power given their status as boss in firms. Now, the word 'institution' here might get us into trouble. We should avoid reading this term in a too narrow "official" mode, such that we refer only to political or legal institutions. There are broader notions of institutions that are more apt. Andrei Marmor, for instance, distinguishes between *conventional* practices, such as forms of art, conventional games, and social ceremonies (which are plausibly not homes for power abuse), and institutional practices, such as legal, political, religious, educational institutions.[11] We should allow that the practices in question here can cover institutional practices in the broader sense used by Marmor. But we should be careful to interpret even this broad notion as broadly as possible: surely medical institutions, theaters, sports leagues, and so forth should count as institutions of this sort. For instance, one might imagine a boxer and his trainer.[12] Here, there is no *office* per se, but there remains a form of institution engaged in by the trainer and boxer—namely, the institution of prizefighting. Given the nature of this practice, however, the trainer maintains substantial power to, e.g., influence and manipulate the good of the boxer—power that can plausibly be abused (say, by insisting he throw a fight for personal gain).

To avoid confusion, therefore, when it comes to institutions in our understanding of abuse of power, I will use the term 'practice' to identify Marmor's sense of institutional practices in the broadest sense. Thus, for the purposes of analyzing power abuse, I will make use of the notion of *practice-based power*. What does this mean? I will call a power relation between *X* and *Y* practice-based so long as the power maintained by *X* over *Y* is explained by *X*'s and

10 One might meaningfully suggest that the forms of power possessed by legal and political officials is variable in just the same way that H. L. A. Hart notes in his infamous argument against John Austin's command theory of law. See Hart, *The Concept of Law*, ch. 1.

11 Marmor, *Social Conventions*, 35–36.

12 Miller, "Corruption."

*Y*'s participation in a particular institution (in the broad sense I mean here). One note on the use of 'explained' here. If I am the president, I will have a number of specific powers granted me by the practice of the federal government and the nature of its constitution. However, these are not the only powers that I have that are *explained* by my being president. I will, for instance, have the power to compel behavior from subordinates, to influence the actions of other individuals even outside my "official" powers. The existence of these powers, while not an aspect of the practice of the presidency, is nevertheless explained by the practice and my position within it, and hence, these powers are liable to abuse. We can now more usefully define power:

> *Power*: For the purposes of investigating power abuse, *X* has power over *Y* to the extent that *X* maintains a form of practice-based power over *Y*. In other words, an essential aspect of the explanation of the power that *X* has over *Y* is a social practice in which *X* and *Y* are participants.

In coming to understand the nature of power liable for abuse, there are a number of tangential notions that we should distinguish from the central notion of practice-derived power. First is *consent*. Power that can be abused (in the sense I mean here) need not be the result of contractual or even consensual arrangements, though of course some will be.[13] The power that a judge possesses over a convicted defendant during sentencing is clear, but it is not, or at least not generally, conceived of as consensual. As a military conscript, I may be assigned a superior officer, but *ex hypothesi*, neither my military service nor my status as subordinate to this officer is consensual on my part. Furthermore, I may be a military conscript and assigned *subordinates*—in this case I may possess power liable to abuse through no consent of my own.

Second is the issue of *conferral*. Of course, all power must be conferred to some extent (i.e., we are not born with it), but the nature and manner of such conferral (whether it is institutional, formal, informal, or so on) is not significant so long as the power is explained by a standing social practice in which the power holder participates.

Third is the notion of *recognition*. Abuse-liable power relations need not be explicit or recognized by the participants involved. For instance, the rules of baseball indicate that the umpire has the power to call balls and strikes. This power, however, grants the umpire the influence noted above: the ability to affect the aims, ends, well-being of the players, and so forth. This power can be abused by, say, refusing to call a pitch a strike until the pitcher pays a bribe. In addition, a judge's power over the accused is relatively clearly demarcated

13 Thanks to two anonymous reviewers.

given the practice of law, but the judge possesses practice-given power not just over the accused but over others as well: community members, legislators (whose aim is to see the law applied), plaintiffs, victims, and so on. And this power can be abused.

Thus, the central idea in understanding the nature of power in an analysis of the abuse of power is that power of this sort is *practice derived*. Many practices will differ (along lines of consent/recognition/etc.), but abuse of power seems to make essential reference to such practices where it arises.

## 3. Abuse of Power, Part Two: Abuse

So what does to mean to abuse practice-derived power? Clearly it is to engage in a kind of behavior that is somehow, well, abusive. But how do we understand this notion?[14] I will consider three possibilities here that I regard as mistaken and argue in favor of a fourth.

### *3.1. The Kantian Framework*

Kant's moral framework is a natural starting place in understanding abuse of power. After all, we may think that what is involved in the abuse of power is a kind of subversion of another's rationality or autonomy, given the exercise of such power.[15] When I refuse to grant you promised humanitarian aid unless you dig up dirt on my political opponent, we may say that I have subverted your ends, compelled you to act in a way that is contrary to your rational concerns; in short, I have used you *merely as a means*. (Perhaps, following Bernard Williams, we might call this the sub-Kantian model.)[16] In a nutshell:

> *X* abuses power over *Y* to the extent that *X* uses practice-derived power over *Y* in a way that is contrary to *Y*'s rational ends.

This proposal, as stated, may explain why we believe that the president abused his power in manipulating a dependent ally for political gain or that an employer, in extorting sexual favors and an attorney sapping a client's fortune are abusing

14 Adrian Vermeule suggests that there are two understandings of abuse of power: "Abuse may be defined in legal terms as action that flagrantly transgresses the bounds of constitutional or statutory authorization, or in welfare-economic terms as action that produces welfare losses—either because officials have ill-formed beliefs or because they act with self-interested motivations" ("Optimal Abuse of Power," 675). However, I argue here that neither of those accounts fits a proper analysis of power abuse.

15 Kant, *Groundwork of the Metaphysics of Morals*, 4:427–29.

16 Williams, "Internal and External Reasons," 78.

their power.[17] After all, these seem to be instances of practice-derived power used to subvert the rational ends of those over whom power is possessed.

However, the sub-Kantian model is not acceptable as an analysis of what it means to abuse power. First, it appears to be too broad. Imagine a judge sentencing someone for a minor crime. The law states that the judge is required to deliver a sentence of three years' hard time. Assuming the judge issues this sentence, she clearly uses her power contrary to the rational ends of the person before her. But this is not an abuse of power. Of course, one might say that such an action is not contrary to the rational aims of the criminal.[18] In a (surprise, surprise) quasi-(sub-)Kantian vein, one might say that if one wills the crime, one wills the punishment for the crime alongside. And while this may be true, it does not solve the problem—surely there are limits of basic justice concerning what the criminal can be said to will. But imagine that the law in this case is extremely unjust: three years of hard time is the sentence for, say, shoplifting a pack of gum. Even if this were so, the judge—though she furthers an unjust system of law—does not abuse power in so sentencing the criminal.

Even if you disagree with my intuitions in this case, it brings up an important set of questions for the sub-Kantian model. In particular, it is hard to see how we should understand the notion of rational aims. For instance, as an employee, I might be required to do all sorts of tasks I find unpleasant. As chair of a department, I may be tasked by the dean with making substantial budget cuts, which would require me to fire a number of graduate students. This is contrary to my rational aims but not an abuse of power on the part of the dean. Of course, one might hold that, in rationally agreeing to take over the role as chair, I have agreed to perform all sorts of tasks I may find unpleasant. But which have I agreed to take on that are contrary to my rational aims and which have I not? It is hard to answer this question without a *further* analysis of when the dean does and does not abuse her power *qua* dean.

### *3.2. The Self-Interested Framework*

A further thing that paradigmatic examples of power abuse seem to share is that, in abusing power, power abusers do so for their own ends or interests.[19] A boss requiring an underling to embezzle money for him uses the power of employer over employee for his own benefit. A president requesting dirt on a political rival extorts information that will benefit the president in a reelection

17 Something like this is suggested by Vrendenburg and Brender. On their view, the abuse of power is in part constituted by acts that "manifest disrespect for a subordinate's dignity" ("The Hierarchical Abuse of Power in Work Organizations," 1339).

18 Morris, "Persons and Punishment," esp. 490.

19 See Vermeule, "Optimal Abuse of Power," 675.

bid or a difficult legislative negotiation. And so one possibility might be that when someone has power over another, the power holder abuses that power to the extent that this power is used in the self-interest of the power holder. This would, for instance, avoid the conclusion that a judge acting in accordance with unjust law abuses power.

To this proposal, we must add a caveat. If I am your boss, and I exercise my power over you in a perfectly upright manner, it may still be that I use this power for the sake of my self-interest. I do so because in so doing I demonstrate competence at my job, which is in my self-interest when it comes to, for instance, staying employed myself. But the fix is easy. We might say that abuses of power are those exercises of power that benefit *solely* the power holder. In the case of the perfectly upright boss, while my actions clearly benefit myself, they do not simply benefit me but presumably benefit many others as well. With that in mind, we might consider:

> *X* abuses power over *Y* to the extent that *X* uses practice-derived power over *Y* for the sake of solely benefiting *X*.

However, this cannot be correct. One can clearly abuse power on behalf of others. For instance, it would be no less of an abuse of power if an employer required an employee who is seeking a satisfactory performance review to embezzle money for the sake of someone the employer knows rather than for himself. Furthermore, one might imagine that the president of country *A* has a good relationship with the prime minister of country *B*, and the prime minister is politically imperiled. The president of *A* then requires a dependent ally to dig up dirt on the political rival of this prime minister, with the aim of keeping the prime minister in power. This is an abuse of power, even though it is not solely for the benefit of the president.

An alternative might be proposed along the following lines. Also inspired by Kant, one might say that power abusers use power in a way that solely advances their particular or *private* ends (rather than their own *interests*, per se).[20] The president who benefits his prime minister buddy does not solely benefit himself, but he does advance *his own ends*. However, this account seems to me far too thin. In particular, we might ask: His own ends *as compared to what*? Imagine that a corporation founder's ends are to become the world's most significant manufacturer of widgets. His employees are in it only for the money. Is acting in a way that advances his company's widget production a "private" end? If so, the founder would appear to be abusing power in so doing, which is clearly absurd. If not, then what is a private end? I submit that to understand this, we

20 See Kant, *The Metaphysics of Morals*, 6:311–6:338.

first need to understand the notion of a *proper* end of a given office, which just seems to me to be asking the question we are here to answer—namely, what is a proper use of practice-derived power versus an abusive one?

### *3.3. The Moralized Framework*

One obvious possibility is that to abuse one's power is to use one's power in ways that are morally illegitimate or unjust.[21] So, we might consider:

> *X* abuses power over *Y* to the extent that *X*'s use of such power over *Y* is morally illegitimate or unjust.

This would rule out the president's untoward pressure on a dependent ally: clearly, doing so is morally illegitimate or unjust. It would also seem to rule out the selling of a vacant senate seat to the highest bidder or bankrupting one's helpless client.

But I think the moralized approach to the abuse of power is both over- and underinclusive. To begin, we might imagine a senator using the power of office to push for a highly inegalitarian agenda—massive tax cuts for the rich at the expense of the poor. While we may think that this is wrongful (morally reprehensible and extremely unjust), it is not an abuse of power.

Furthermore, it is not at all clear that using one's power in morally legitimate ways cannot constitute an abuse. Imagine, for instance, that an employer believes very deeply in a morally important cause, so much so that she threatens to withhold satisfactory performance reviews from those employees who fail to donate substantial sums to that cause. This case strikes me as an abuse of power, but it is not at all clear that the action is or must be morally illegitimate, at least independently of a consideration of the significance of the cause. To put this in concrete terms, imagine that the employee is a medical skeptic who refuses to submit his child to a routine medical procedure that would save the child's life. Without intervention, the child will die. The employer knows this and threatens to withhold a satisfactory performance evaluation from the employee unless the employee takes his child to the doctor for this routine

21 Something like the moralized framework is suggested by Raday, according to whom abuse of power is mainly treated as the use of power contrary to individual human rights ("Privatising Human Rights and the Abuse of Power"). In addition, Jaggar and Tobin also suggest that "power and vulnerability are abused when people take wrongful advantage of them" ("Situating Moral Justification," 388). Though not discussing abuse of power directly, Miller suggests that corruption is essentially a form of immorality: "Corruption is . . . one species of immorality." He continues: "corruption in general, including institutional corruption frequently, if not typically, involves the despoiling of the moral character of persons and in particular, in the case of institutional corruption, the despoiling of the moral character of institutional role occupants qua institutional role occupants" ("Corruption").

procedure. I find it very implausible to hold that the employer has done something morally wrong here. Indeed, or so I suspect, exercising power in this way may very well be morally required. But I also find it quite plausible to hold that the employer has *abused her power*—has exercised her power in a way that is beyond its appropriate extension.[22]

### *3.4. The Relationship Framework*

The above proposals suffer from a common defect. In each case, they understand the abuse of power with reference to a means of evaluation that is external to the power relation and its practice—whether the behavior of the power holder is, e.g., compatible with the subordinate's rational aims, purely self-interested, or morally inappropriate or unjust. But in considering power abuse, we are typically led to look not to external methods of evaluation but rather to *internal* methods of evaluation—evaluation internal to the practice that governs the power relation.

What is an internal evaluation of this kind? Notice that practice-given power relations do not just spring from the ether. They are established, and they are established *for a reason*—there is a *raison d'être* or *point* of the power relation. For instance, the power the president has over a dependent ally has a *raison d'être* of this kind—namely, the strategic goals of the constituent nations. In the case of an employer/employee relation, the *raison d'être* includes the health of the firm, the efficiency of a particular office, and so on.

Leaving aside some details to be filled in, I think we are in a place to offer an account of the abuse of power that is more successful than previous attempts. To see this, note the example given at the end of the previous section. It seems right that, though her motives are as morally pure as the driven snow, an employer who withholds a satisfactory performance evaluation unless an employee assents to take his child in for a medical procedure uses the power she possesses over this employee beyond its appropriate extension. In my view, then, the *appropriate extension of power* in this case is limited to the exercise of practice-derived power to advance, promote, or contribute to the *raison d'être* of the power relationship. Assuming that the *raison d'être* of the employee/employer relationship concerns the health of the firm, the efficiency of the department, and so forth, the use of power over an employee to compel the employee to do the right thing by his child is beyond the *raison d'être* of the employee/employer relation. And hence, in this case, power is abused.

22 Note that this provides another plausible example of the divergence between *corruption*, on the one hand, and power abuse, on the other. This is an example of power abuse, but the employer here is not corrupt—plausibly given the linkage between corruption and moral degradation, as already noted.

This is the rough idea, but accounting for this basic thought in precise terms is challenging. To see this, consider one possibility:

> *X* abuses power over *Y* to the extent that *X* uses practice-derived power in a way that does not conform to the *raison d'être* of the practice from which such power is derived.

This account is unacceptable because it wrongly characterizes *mistaken* uses of power as *abuses* of power. For instance, I may act, as holder of power of attorney, against the best interests of my client. But I may do so as a simple mistake rather than as a genuine abuse of power.

This reflection seems to indicate that whether or not one is abusing one's power depends, at least in part, on one's own beliefs, attitudes, or intentions. Perhaps intentions are the correct approach: Blagojevich intended to line his pockets; I, the mistaken attorney, intended to promote my client's interests, though I did so incompetently. Perhaps, then:

> *X* abuses power over *Y* to the extent that *X* uses practice-derived power with the intention of promoting states, acts, events, and so on, that do not conform to the *raison d'être* of the practice-derived power relation.

This formulation correctly characterizes Blagojevich as abusing power and the mistaken attorney as not abusing power. This is an improvement, but further refinement is required. Imagine an elected official who makes every decision for the sake of getting reelected. Assume for the purposes of argument that his own reelection is not within the *raison d'être* of this elected official's practice-derived power relation.[23] Nevertheless, let us say that the elected official actually does advance the point of this power relation (by making good decisions, etc., perhaps because he believes doing so is the most efficient way to get reelected). The previous account would seem to categorize this official as abusing power because he uses his practice-derived power *for the sake of* his reelection, which—*arguendo*—does not conform to the power relation's point. This seems overinclusive. In addition, this account is underinclusive. Imagine a delusional fascist leader who genuinely believes that his own self-interest is in the best interest of his nation and embezzles all of the nation's funds to build a gigantic palace in the Alps. This plausibly constitutes an abuse of power even though, given his delusion, the fascist leader may have sincerely done so *for the sake of* the interests of the nation.

23 This, I think, is a contentious assumption in the case of democratically elected officials, but I will entertain it *arguendo*.

What, then, explains our intuitions in the case of the elected official or delusional fascist dictator? I think this: in the case of the elected official, there is at least a reasonably expected correlation between the pursuing reelection and the promotion of the point of the power relation (good governance, etc.). In the case of the fascist leader, however, there is no such correlation: no sensible person would believe that promoting the dictator's own interests is correlated with the interests of the state. So rather than focusing on the *intentions* of the agent, it seems better to focus instead on the agent's *beliefs*, specifically whether the agent can or can reasonably foresee the action they engage in or the states they promote as correlated with the *raison d'être* of their power relation. With that in mind, consider the following adjusted proposal:

> *X* abuses power over *Y* to the extent that *X* uses practice-derived power in a way that is not foreseeably correlated with the promotion of the *raison d'être* of the power relation.

We should understand 'foreseeable' to be indexed to the *agent's* beliefs, with a *weak* test of reasonability (which would presumably rule out the delusional dictator). In the case of the mistaken attorney, he does not foresee that his act would lead to his client's misfortune (though one might imagine that if he is mistaken *enough*, abuse of power could arise in the same way as the fascist leader). In the case of the official seeking to get reelected, it is not the case that his action is not so correlated; typically, we should imagine, officials seeking to get reelected will act competently (or at least as the populace desires). One might wonder how strong the correlation must be in typical cases: must it, say, maximally contribute to the *raison d'être*? Contribute to a sufficient degree? I am tempted to hold the latter, but in a way in which the "sufficient" degree is tied to the particular practice in question. For instance, selling a senate seat to the highest bidder might foreseeably correlate with the point of a state governor's power *to some degree* (i.e., at the very least there is a warm body in the seat), but surely it does not do so to a sufficient degree given the demands of the office. Further refinement of this notion is surely permissible given additional consultation with considered judgment.

Interpreted in this way, the above account seems to get the proper answers. In the case of Blagojevich, he could be expected to know that selling a senate seat to the highest bidder is not the sort of action typically correlated with to the *raison d'être* of his political power relation. Likewise, the *raison d'être* of the president/ally relation is clearly not the political fortunes of the president but rather the intertwining strategic goals of both nations. Plausibly, the *raison d'être* of the relation between a judge and a convicted criminal is the maintenance and application of the rules of law. And so in acting in accordance with that *raison d'être*, the

judge is not abusing power by imposing a sentence, though of course the judge may deliver a manifestly unjust sentence given the nature of the law itself. And an attorney who robs his client is using his power in ways that cannot be foreseen to contribute to the interests of his client. Finally, requiring a medical skeptic to take better care of his children is (at least in virtually every such relation) beyond the *raison d'être* of the power relation that exists between boss and employee and thus cannot be reasonably expected to be correlated with that *raison d'être*.[24]

### 3.5. *Contrasts*

Two notes of contrast would be helpful here. Joseph Raz, in describing the notion of *arbitrary power*, holds that "an act which is the exercise of power is arbitrary only if it was done either with indifference as to whether it will serve the purposes which alone can justify use of that power or with belief that it will not serve them."[25] This may seem superficially similar to the view I have proposed here, but Raz's view is importantly different. First, a minor point: Raz's view would continue to classify the delusional fascist dictator as *not* exercising arbitrary power, when it seems to me clear that this is a case of power abuse. But more importantly, notice that the *raison d'être* of a power relation need not *justify* the use of power. A senator who uses his power to advance immoral or unjust ends acts within the *raison d'être* of the senatorial power dynamic. But this clearly does not justify the *use* of such power—for justification, clearly some reference to the moral quality of the ends is required. Thus, arbitrary power (as Raz conceives of it) differs from the abuse of power in this way. Just because the power I maintain over you is not *justified* does not mean that there is no distinction between abusing my power and not abusing my power.[26]

24 This account also sheds additional light on the difference between power abuse and corruption. The city council member accepting a bribe despite acting as he deems best in no way *uses* practice-derived power in a way that tells against the *raison d'être* of office. There is simply no abuse of power here. But no plausible analysis of corruption would leave out such behavior. Furthermore, the president who seduces an impressionable intern is clearly using his practice-derived power in ways that are not foreseeably correlated with the point of his presidential power, and he knows it. But I hesitate to say that the president is *thereby* corrupt. Of course, there may be a *sense* of corruption that applies to the president in this case: the more general sense of moral degradation (as suggested by Miller, "Corruption"). But this simply adds to the sense in which power abuse, which need not be an instance of moral degradation, and corruption are nonidentical concepts.

25 Raz, "The Rule of Law and Its Virtue," 219.

26 Note also that Larissa Katz suggests a general account of abusing *rights* that is similar to Raz. For instance, she holds that the point of the power of ownership rights is, e.g., the "setting of the agenda" for a particular thing. For Katz, this limits the morally justifiable reasons one has with respect to that thing. See Katz, "Spite and Extortion," 103–7. However, her account of the abuse of this right requires that there be a political justificatory rationale

Second, my proposal should be distinguished from another superficially similar view. Emanuela Ceva and Maria Paola Ferretti hold that political corruption is understood in the following way: "There must be a public official who (1) acts in her institutional capacity as an officeholder (*office condition*) (2) for the pursuit of an agenda whose rationale may not be vindicated as coherent with the terms of the mandate of her power of office (*mandate condition*)."[27] Leaving aside the concern with public office holders (Ceva and Ferretti's topic is, after all, *political* corruption), the *mandate* of public office should be understood as conceptually distinct from the *raison d'être* of that particular power relation or institutional office. Notice that the mandates of office or of a particular power relation will necessarily include the various written rules or norms of behavior. But as I argue in more detail below, the *rules* that govern power use are or can be distinct from the *raison d'être* of a power relation, and hence their breach may not constitute an abuse of power.[28]

Now, one might challenge this proposal and hold instead that it is these very "deontological rules" that govern the exercise of power in an institutional context that are a viable alternative to the relationship framework.[29] After all, these rules are internal to the practice itself just as much as the *raison d'être*. But this is not plausible as a substantive matter. First, one might imagine, say, that an employer is required to give certain reports about employees every thirty days—this is a written rule that governs her interaction with her subordinates and is part and parcel of her "mandate" when it comes to such power. She might decide to break this rule, however—maybe because she knows this rule is a hinderance to rather than a catalyst for worker productivity. This is not plausibly an abuse of power, though it may be a breach of what might be called the deontological rules that govern the role in question.[30]

---

for the office/practice in general, and that abuse of power in that case would be use of that power that runs counter to the political justification of the office.

27 Ceva and Ferretti, *Political Corruption*, 19.

28 In some cases, these rules can form *part* of the *raison d'être* (e.g., in the case of the police, insofar as at least one point of the police power relation is the upholding of the very written laws that apply to the police, or, in the case of the US president, who is sworn to uphold the Constitution, which itself determines the proper extension of presidential power), but in other cases, it need not.

29 Thanks to an anonymous reviewer.

30 An anonymous reviewer responds with the following argument: what constitutes power abuse can change over time. We may consider corporal punishment an abuse of power now, but the power relation of school principal–student still retains its *raison d'être* even during times when corporal punishment was not considered an abuse of power. Hence it must be the *rules*—that now, but not then, forbid corporal punishment—that define abuse of power, not the point or *telos* of the power relation (which would never, I assume, have

Summing up, on my view, abuse of power is a fundamentally contextual notion. Whether someone has abused power clearly depends on the nature, function, and aims of the practice-determined power relationship they occupy. It is perfectly possible for morally upright individuals who never act contrary to moral principles to abuse power if the relationships they inhabit lack moral principles as a guiding *raison d'être*. Furthermore, it is possible for individuals who act in morally egregious ways to nonetheless fail to abuse their power given the nature of the relationships they inhabit.

### *3.6. Practices, Power, and Raisons d'être*

A key concept in understanding the abuse of power is a power relation's *raison d'être*. But what is this? And how do we determine it? Above, I suggested that the *raison d'être* of the president/dependent ally relationship is, e.g., the advancement of national strategic interests in the region; that the *raison d'être* of the judge/convicted relation is the maintenance and application of the rules of law; and so on. But why should this be? Is there a general account of the nature of a *raison d'être* of any particular power relation that would bear out the relevant claims and/or answer our questions concerning what the *raison d'être* of any given such relation is?

Consider first the structure of social practices in which such power relations are embedded. Such practices have both a constitutive set of *rules* (i.e., what it is to be engaged in such a practice, the "deontological rules," the standards that constitute holding office, and so on) and also a constitutive *aim* or *reason* for the existence of the given practice.[31] This reason or aim itself can be conventional.[32] It could also respond to some general human need, specific coordination prob-

permitted corporal punishment). But this analysis is mistaken. The relationship framework requires that the person with power (the principal) reasonably foresee that their action is correlated with the promotion of the point (i.e., education of children) of their practice (i.e., schooling)-derived power. And I think that has been true of most principals, even in the past; generally, people reasonably thought, given evidence and tradition, and so on, that corporal punishment was an acceptable means to promote educational outcomes and good behavior. But we now know that it does not, and this knowledge is widespread; hence, school disciplinarians cannot reasonably foresee that it will. And hence, now corporal punishment of children is an abuse of power, whereas it was not at other times. The change in deontological rules of the institution is therefore an *indicator* rather than a proper explanatory principle in the change in the status of corporal punishment *qua* power abuse.

31 See Marmor, *Social Conventions*, 5.

32 This is what Marmor calls a "deep convention" as opposed to a "surface convention." The surface convention of, e.g., not wearing white to a wedding is given by the fundamental (deep conventional) aim of showing respect to those who are to be married. See Marmor, *Social Conventions*, 74–76.

lem, or some other concern.[33] But in service of this general aim, some practices give rise to power relations that have a constitutive aim or end *given* by the practice. To see how this might bear out, take baseball. The umpire has power over the pitcher, batter, etc.: power to call balls and strikes. But what is the aim of this power relation? In this case, the aim is to have a rule-guided authority on balls and strikes. That there is such a need is dependent on the practice itself: one cannot play a baseball game (or in any event play it effectively or efficiently) without some such rule-guided authority. And it is *this* aim that, in my view, constitutes the *raison d'être* of a given power relation.

With this in mind, there are (at least) five important notions from which the *raison d'être* of a power relation must be distinguished. First is the *motives* of the individual participants to a practice. Take the baseball game. One might imagine that the umpire just wants a payday; the pitcher wants a strikeout to advance his career. But the aim of this power relation refers not to the aims of either participant but rather to the function or role this power relation plays within the practice—namely, the need for a rule-guided authority. The second notion that must be distinguished is the *beliefs* of the individual participants about the *raison d'être* of the power relation. I may be an employee at a firm and believe that the whole point of such employment is to make *me* money. But it is still the case that the *raison d'être* of the employer/employee relationship is at least in part characterized in terms of the health of the firm, the advancement of the firm's ends, and so on. The third notion is *other domains of evaluation*. There is no guarantee that the *raison d'être* of any particular power relationship of this kind will overlap with other domains of evaluation, such as morality or practical normativity generally. It depends on the relation and its place in the practice in which it is embedded.[34] Fourth, we must consider the *behavior*, even *tolerated behavior*, of those in power. Imagine a Hollywood producer of the bad old days who simply treats the casting couch as his sexual breeding ground. Indeed, such behavior from such individuals may be extremely widespread and tolerated, but this does not entail that the point of this power, derived as it is from the practice of making movies, is the sexual exploitation of ingenues. The fifth thing that must be distinguished from the *raison d'être* is the so-called *code of conduct* of a given power relation. Such codes or rules will, if they are

33 See Lewis, *Convention*.

34 Note that one person's power over another can be embedded within multiple different practices. For instance, a father could have power over his son by virtue of the parent/child practice but also by virtue of the fact that the son is his father's apprentice. But this causes no problem for the analysis: in this case, as this power relation will have different *raison d'être* given the different practices in which it is embedded. The father could potentially abuse his power *qua* parent but not *qua* mentor, and so forth.

well designed, be such that conformity to them will *in typical cases* lead to a furtherance of the constitutive end of the relation. And so in well designed codes or norms, there will be a strong correlation between norm-conforming action and power nonabuse. But this need not always be the case. If, for instance, the code of conduct is known to be badly designed and foreseeably lead to circumventing the *raison d'être* of a power relation, it could be that conforming to such a code *is* an abuse of power and that circumventing it is not.[35]

Putting this all together, then, the *raison d'être* of a practice-given power relation is the aim or end of the power relation *given* the practice in which it is embedded. Consider now the paradigmatic examples we have so far been discussing. Rod Blagojevich maintained a certain power over his constituents, namely, the powers held by a governor (including, for instance, the power to appoint a senator). What is the reason, given the practice (here, the practice of state government), that such a power relation exists? The answer (plausibly) is the efficient exercise of state government in accordance with the Constitution of Illinois. Similarly, the *raison d'être* of the power relation between employer and employee is, as already suggested, the furtherance of the goals of the firm, the efficient running of a department, and so on. But the use of such practice-given power to further other goals—even morally appropriate ones—when those goals are foreseeably uncorrelated with the *raison d'être*, constitutes an abuse of power. The power relation between a president and dependent ally is embedded within a certain practice of international diplomacy, and the aim of such a power relation is the furtherance of the strategic goals of both nations. It is not the digging up of political dirt.

### *3.7. Is the Relationship Framework Overinclusive?*

An important objection to the analysis of abuse of power that I have suggested here is that it is overinclusive: it categorizes what may seem intuitively not to be an abuse of power as an abuse of power. Consider the following case:

> Jennifer has a young daughter, Emily, who, by conventional standards, is extremely cute. Because of this, she enters Emily into grueling beauty pageants and forces her to work long hours in front of cameras for photo

35 It is also worth noting that *raisons d'être* are not set in stone but can morph and change with the nature and content of the social practice. For instance, it may be that the *raison d'être* of social practices like marriage, childrearing, even political power and office, has changed over millennia (and may change again), giving rise to new understandings of what does and does not constitute an abuse of power from within such practices. (Of course, when and how such *raisons d'être* change will be subject to indeterminacy and grey area given the nature and inertia of social institutions.) I take this to be a simple consequence of the nature of social practices. Thanks to two anonymous reviewers for helpfully making this point.

> shoots and local commercials. Rather than saving the child's earnings, Jennifer spends it on her own lavish lifestyle.

Clearly, Jennifer is doing something wrong. But is this an abuse of power? The relationship framework as I outline it here would seem to say yes. After all, Jennifer and her daughter are in a power relationship (parent/child) within a given practice (i.e., the practice of family life, raising children, etc.). Furthermore, we can assume that the *raison d'être* of this power relation is certainly *not* funding a lavish lifestyle for the parent but rather the care and flourishing of the child. But this may seem strange. Why think that a parent's abuse of her child constitutes an abuse *of power*? Why is this not just a case of irresponsible parenting?

I am not compelled by this argument. Nothing stops us from saying that in addition to abusing her power as a parent, she is *also* an irresponsible parent or *also* behaving immorally. Indeed, it may sound a bit strange to say that Jennifer commits an abuse of power, but this may be explained by noting that, plausibly, the more significant violation is failing to be a responsible parent to her child, exploiting her child for her own gain, and so on. Indeed, there seem to be cases that are quite clearly instances of parent/child abuse of power. To see this, consider:

> Ronald Crump is a self-aggrandizing politician with a wickedly inflated ego. To stroke his confidence, he insists that his children, over whom he wields great power, introduce him at campaign events with over-the-top stories of his prowess as a father and family man.

This seems to me a clear abuse of power—an abuse of Ronald's power over his children. If this is right, I am inclined to hold that the parent/child relationship is one that can play host to abuse of power, even if other, more significant violations have a tendency to drown out the accusation of power abuse.

To see the next objection, consider:

> Joe is a midlevel mafia enforcer with a number of underlings. One day, troubled by the increasing violence of his job but without the courage to do it himself, he goes to one of his henchmen, Bob, and tells him that unless Bob becomes an informant for the FBI and turns in the head honchos of the crime organization, Joe will tell the Big Cheese that Bob has been skimming the take. Bob complies, informs, and the syndicate is disbanded.

In this case, we are to imagine that Joe has power over Bob that is embedded within a particular social practice, namely, an organized crime syndicate. It seems quite obviously not correlated with the *raison d'être* of this power

relationship (characterized, as it is, by, e.g., shaking down vulnerable business owners or threatening rival gangs) for Joe to insist that Bob become an FBI informant. And hence, or so it would seem, my proposal seems to characterize Joe's actions as an abuse of power.

But, again, this may seem overinclusive. Why think that, in forcing one of his underlings to do the right thing and help destroy the organization, Joe's action constitutes an abuse of power? Ultimately, I think we should, in fact, accept this conclusion. Part of my temptation here involves the cost of alternatives. For instance, one might suggest that the current view could be amended by holding that for an abuse of power to occur, it must be the case that the *raison d'être* of the power relationship itself has moral value. If we say this, then we could pretty straightforwardly suggest that Joe's actions do not constitute an abuse of power because the power relationship that he maintains with Bob has a morally unacceptable *raison d'être* (namely, the furthering of the interests of an organized crime syndicate). But we should reject this amendment. It would imply that no abuse of power can occur unless the *raison d'être* of the power relationship has moral content. But this is clearly wrong. Indeed, it seems wrong in the mafia case: if a mob boss exploits his power over an underling for sexual favors, this seems like a clear case of power abuse.

Rather, I think what drives our reaction to Joe's case is not so much that no abuse of power can occur within a mafia framework but rather a concern about what precisely Joe is asking Bob to do. After all, it seems like the "right thing" or at least a *good* thing. And though the right thing is not foreseeably correlated with the *raison d'être* of the mafia organization and its internal power structures, it seems plausible to hold that someone within that structure is not misusing his power when he forces an underling to become an FBI informant.

Now we have a choice. We could amend the relationship framework to hold that abuse of power does not occur when an would-be abuser does the right thing. But this proposal is too strong: it would have the effect of holding that the employer of the medical skeptic does not abuse employment power over her employee. The other possibility is better. We should not try to shoehorn the plausible judgment that Joe's use of power over Bob had moral content or was the right thing to do into the analysis of whether or not Joe abused this power. Instead, we should treat these questions—*Was $\phi$ an abuse of power?* and *Was $\phi$ the right thing to do?*—as separate inquiries, at least for the purposes of analysis. If squeamishness at describing Joe's case as abuse of power is, as I submit it is, down to the fact that Joe does the right thing or something for which there was strong justificatory reason, we should treat this as shedding light not on our analysis of the concept of power abuse but rather on the *normative significance* of power abuse. To this I now turn.

## 4. NORMATIVITY AND THE ABUSE OF POWER

If my proposal is correct, there arises a serious question about the normative significance of power abuse as a category of action evaluation. After all, Joe abuses his power, but not in a way that is wrong or normatively ruled out. How, then, should we understand the *normative* consequences of the abuse of power?

We should, at the very least, reject the claim that one is required to avoid abusing power. Consider again the case of the medical skeptic. Imagine that the skeptic's boss notices that several of her employee's children will likely die of a very serious and painful illness if not given an utterly trivial medical treatment. If there are no other options available, and the employer abuses power for the sake of protecting those children, then it seems right to say that the employee acted in a way that was, on the whole, permissible. If that is right, then the mere fact that we abuse power does not entail that we act wrongly.

However, there does seem to be at least some reason to avoid power abuse. Imagine in the case of the medical skeptic that the employer has two options to save the lives of her employees' children. The first is to abuse her power and force the medical skeptic to provide the medical treatment to his children if he is to keep his job. The second is to offer to personally pay him $2,000 to do so. It seems right in this case that the employer faces stronger reason to pay her employee rather than to withhold a positive employment evaluation. There is something about the employer abusing her power, in this case, that seems normatively unsavory *in comparison* to the other alternative. This seems generally true. It would be better for, say, Joe to convince Bob through non-power-abuse means to become an FBI informant. (Perhaps he could simply buy Bob a beer and explain, or promise Bob a big payout, or employ some other means of persuasion.) It would be better for the president to find some other way of dishing dirt on his political rival rather than abusing his power to do so, leaving all other things equal. When other means are available to accomplish the same end, abusing one's power seems normatively disfavored.

But one challenge to this proposal concerns what the reason is to avoid power abuse. The challenge here arises from the thought that the abuse of power is a very diverse category, given, as noted already, the diversity of practices and the power relations so embedded.[36] It is not obviously the case that all abuses of power share some feature that is normatively significant and tells

36 Perhaps the most obvious possibility is that, e.g., in abusing her power, the employer is *coercing* her medical skeptic employee into action. But this cannot be the whole story. After all, not all instances of the abuse of power are coercive—see the abusive attorney who drains the bank account of his comatose client or the governor who sells a senate seat to the highest bidder. Given the diverse ways in which individuals possess power, let alone

against abusing power. However, this should not unduly concern us: as I shall now argue, the fact that one abuses power in $\phi$-ing is *itself* a reason not to $\phi$.

## 5. ABUSE OF POWER AS A REASON

I argue that the fact that *X* abuses power over *Y* is a reason against so doing. My defense of this claim relies on the following principle:

> *Power Expectations Principle*: In any practice-embedded power relation, a subordinate *Y* can adopt a rational normative expectation that power-holder *X* will confine the use of *X*'s practice-derived power over *Y* to instances in which the use of such power foreseeably conforms to the *raison d'être* of this power relation. This rational expectation is (or implies) a reason for *X* to confine such power to uses that so conform.[37]

The Power Expectations Principle is really a conjunction of two claims, the first concerning what a subordinate can normatively expect when in the midst of an embedded power relation, the second concerning the normative significance of such an expectation.

A word on expectations. The word 'expectation' should be disambiguated between what might be called *predictive expectations* (e.g., when I say to a fellow bus rider, "I expect it to rain") and *normative expectations* (e.g., when a parent says to their child before a fancy dinner, "I expect you to say 'please' when asking for the salt"). I use the notion of expectation in the normative, not predictive, sense.[38] And there are two features of the notion of a normative expectation that I mean to bring out. First, normative expectations can and will carry with them negative reactive attitudes if they are thwarted. (This is in part what makes them "normative.") But second, some such expectations will be *rational* and others *irrational*—alternatively, one might call them *fitting* or *unfitting*. One rational normative expectation, for instance, would be for my employer to pay me according to our agreed wage when I have completed my work rather than threatening to withhold such payments for personal ends. And some expectations of this kind may be irrational—e.g., I may expect my

abuse it, it seems plausible to hold that the fact of abuse of power *is itself* of normative significance.

37 The qualifier 'imply' is intended to render the power expectations principle ecumenical between views according to which the underlying practice is normative via the fitting expectation and those views according to which the underlying practice is a reason and the fitting expectation is merely *indicative* of the underlying reason.

38 It could be, for instance, that I know you are corrupt or sexually exploitative, and so forth, and I know full well that you will abuse your power.

employer to pay for my children's schooling or to continue to pay me even if I fail to show up or consistently show up drunk, and so forth.

I argue that these rational expectations have normative significance below, but I will start here with the first claim embedded in the Power Expectations Principle, namely, that subordinates can adopt a rational normative expectation that those who hold practice-derived power over them will confine the use of such power to the *raison d'être* of the power relation. Why believe it? I will offer two arguments here. First, it is plausible on its face. When I am a subordinate in a power relation, and I understand the nature and function of this power relation (that is, I understand the *raison d'être* given the practice in which we participate), I expect that you will restrain your use of power over me to those cases that are at least foreseeably correlated with the point of your power. Of course, it may be that power holders placed within such power relationships do in fact go beyond the *raison d'être* of such power relations. (One might imagine, say, rampant sexual harassment of employees in certain companies.) But even if this is true as an empirical matter, it remains the case that I have a fitting *normative* expectation that if you have power over me, you will (foreseeably) *stick to the point*.

Second, this claim seems to make sense of everyday experience. Take a humdrum example. Imagine that you are an usher at a professional baseball game, and I present you my ticket to show that I belong in the section over which you have authority. This relation is one of power—given the practice in which we are engaged (the practice of, e.g., spectator sports), you as an usher have power to dismiss me from the stadium, seat me properly, etc. But in this power relation, I, as a spectator with a ticket, have a normative expectation that you will use this power only in the proper way, e.g., seating me upon presentation of a valid ticket, not that you will seat me only if I present a bribe or agree to donate to Oxfam International or agree not to speak sharply to my kids when I get home. Note that not every possible expectation on the part of the person over whom power is wielded will generate practical reasons for the power wielder. I may expect that if I present to you a valid ticket, you will not just seat me but also give me $100. This may be a normative expectation, but it is certainly inappropriate in such a case. But why? What distinguishes the expectation that you will seat me and the expectation that you will cough up? Plausibly, the answer is that giving me one hundred dollars does not have anything to do with the point of your having this power over me.[39] It seems right to say that if we are both

39 A reader has suggested that this expectation may have more to do with the contractual obligations of the usher. But this misses the point: it is the normative relation between the ticket holder and the usher that is doing the work here, not the normative relation between the stadium owner and the usher. But if this is bothersome, assume that the usher

participating in a given practice, it seems fitting for me to expect you to limit your uses of practice-given power to the *raison d'être* of the power relation, given that practice. The relation of power between, e.g., ushers and spectators appears to be that paying customers are properly seated, making sure everyone has an enjoyable time at the ballpark, maintaining order, and so on.

Objection: one might argue that the first part of the Power Expectations Principle is overly broad. Imagine I am unjustly imprisoned in a jail cell. The warden has power over me—to release me, keep me locked up, and so on. But it would seem that given that the *raison d'être* of the warden–prisoner power relation is to (in part) keep prisoners imprisoned given the duly applied dictates of the law, it would appear that I, the prisoner, could rationally expect that the warden to, e.g., keep me locked up. But this is absurd, and a general point might hold: in cases in which a particular practice suggests that the point of a power relation is to, say, harm me, treat me immorally and unjustly, and so on, surely it would not be fitting for me to expect that I be treated so (especially if I have been thrust into such a power relation without choice)!

I disagree. This objection is based on one (or more) of four potential confusions. First and most importantly, it is simply incorrect to say that the *raison d'être* of the power relation will not form a fitting expectation of the warden on my behalf. After all, if the warden proposes not simply imprisoning me unjustly but also subjecting me to daily beatings, humiliation, or personally motivated abuse, I will *certainly* expect that the warden not do so. And while I may form this expectation on many grounds (including its profound immorality), surely *one* basis of this expectation is that *this is not how our power relation is supposed to work*. So clearly, the *raison d'être* will form a locus of expectation at least to this degree. Now, of course, I may also (quite fittingly) expect that the warden release me from my unjust imprisonment. But second, this perfectly sensible normative expectation is fully compatible with the Power Expectations Principle. Nothing in that principle holds that it is the *sole* source of normative expectations. If I am imprisoned unjustly, for instance, I can form normative expectations based on my unjust treatment. I can form expectations that the warden act for my benefit rather than my ill insofar as he clearly has reasons to do so. But none of this is a violation of the Power Expectations Principle. Third, the Power Expectations Principle says nothing about whether I can or cannot fittingly expect that you *use your power*; rather, rather it says only that I can rationally expect that the use of your power, if you are going to use it, is confined to the point of our power relation. So in the prison case, I may be

---

is simply a volunteer and maintains no contractual obligation whatever. Nothing about the case seems to change.

unjustly imprisoned and have no rational expectation that you will keep me here. But I can rationally expect that, insofar as you are going to use your prison warden power over me, you confine it to the point of our relation rather than, e.g., using me for your own ends. Fourth, recall that the Power Expectations Principle holds that it is fitting to attribute reasons to the warden to stick to the point of our power relation, not that I *will* do so. Imagine, for instance, that it is part of my aims to remain imprisoned—I see myself deserving of punishment and believe that the long arm of the law got me in a fair cop. If the warden uses his practice-derived power to release me early, it would seem perfectly rational for me to complain specifically on grounds of the nature of our power relation: the duly applied dictates of law. Of course, most will not complain under these circumstances. But those who do would do so fittingly, just as they would fittingly complain if the warden used his power to force the prisoner to engage in humiliating activities for the warden's amusement.

So far, I have argued in favor of the first part of the Power Expectations Principle—namely, that in any power relation, subordinate *Y* can adopt a rational normative expectation of power-holder *X* to constrain the use of practice-given power in ways that conform to the *raison d'être* of the power relation. But the second part of the Power Expectations Principle also requires defense. Though some might, I do not hold that a rational normative expectation to act in some way entails a reason to do so.[40] But even if this entailment does not hold in the general case, it seems to hold in the context of power relations. And the explanation concerns quite general facts concerning the normative expectations of those over whom one has any kind of power.

Note that every person over whom power—the ability to influence well-being, interests, behavior, and so forth, whether practice-derived or not—is wielded is thereby put in a position of vulnerability. This in and of itself is not a bad thing. We recognize that power over others and others' power over us necessary features of many aspects of our lives. But this vulnerability plausibly generates reasons for those who wield power. This position of vulnerability makes the power holder specially *answerable* to the normative expectations of the vulnerable in that power relation. If I have power over you, I am answerable to *your* normative expectations (or perhaps, put it more precisely, to the normative expectations that would be fitting for you to adopt). Now, I may be answerable to the normative expectations of others as well, but because your interests, your actions, the satisfaction of your aims, and so on are susceptible to my influence, this seems to put special weight on my responsibility to take your normative expectations seriously. Note that this is a weak principle. The

40 See Darwall, *The Second-Person Standpoint*.

normative expectations you have of me do not attribute to me an obligation to act come what may—they attribute to me only a *reason*, a *consideration I must take seriously*. There may yet be other factors—including, importantly, moral factors—that determine how I am to use my power. The normative expectations hold only that I am answerable to you in the sense that, if I do not conform to your expectation of me, there had better be some justificatory reason.

One major objection remains. Just as in the case of fitting expectation, one may be concerned that there is no *reason* to conform to the *raison d'être* of a given power relation (even in the presence of such fitting expectations) in the context of power relations that seem morally heinous. But *in general*, the Power Expectations Principle does not deny this. There is surely no reason whatsoever for someone to, e.g., engage in the practice of slavery or to use the power derived from or explained by that practice. Rather, the Power Expectations Principle holds only that *if such power is to be used*, it not go beyond the point of the practice. (If, for instance, I repent and decide to allow my slaves to escape or to help them become free people, this is not the *use* of practice-derived power but rather an *abdication* of such power.) But (hence the qualifier 'in general') perhaps there are cases of morally bad power relations, in which the morally optimal option involves power abuse. Does this show that the principle is incorrect? Does it show that there is no reason to refrain from abusing power?

The answer is no. This is because even if it would be normatively preferable for a power holder to abuse power in order to bring about the morally correct outcome (surely right), there remain reasons for the power holder not to abuse power to *other* ends. To see what I mean here, consider a case in which a power holder in an oppressive regime acts beyond the purview of his power to, e.g., humiliate his subordinates. The subordinates will clearly have a fitting moral complaint. But they will also have a fitting complaint given his role in the oppressive regime. Whatever else you might say about it, their relationship *is not supposed to work that way*. And if that is right, we should accept the normative significance of the *raison d'être* of even morally outrageous power relations, even if (in some cases) abusing power is the right thing to do, all things considered. Now, one might say that the normative significance of the way this relationship is supposed to work is limited only to particular cases (i.e., humiliation, yes; freeing from oppression, no). But recall that there are good reasons to be skeptical of the suggestion that there exists "no reason" to perform some action in cases in which, if there were such a reason, it would be massively outweighed by other considerations, as they clearly would be in the case of freeing someone from oppression or slavery.[41] It may seem that I

41 Schroeder, *Slaves of the Passions*, ch. 6.

have no reason to, say, be an authoritarian prison warden in an abusive penal system rather than abusing my power to benefit the prisoners. But this would be a mistake: there is such a reason, but in such cases, the reason is so massively outweighed that it is of slight account.

If all this is right, then abuse of power is itself a normatively significant fact—abuse of power is a failure to conform to the rational normative expectations of those over whom power is exercised.

## 6. Conclusion

Abuse of power seems to be one of the most significant categories of misdeed in public life. After all, presidents do not get impeached for immorality, imprudence, injustice, or a failure of virtue. They do get impeached, however, for abuse of power.

But as I have so far indicated, it is not so easy to understand just what abuse of power is and why we may have reason to avoid it. I have argued here that the proper understanding of abuse of power is to be found from *within* practice-embedded power relations (i.e., in the *raisons d'être* of power relations), not from *without* (i.e., not in violations of moral norms, norms of justice, or advancement of one's self-interest). Furthermore, I have argued that there is good reason to believe that an account of this kind can deliver on the claim that we have practical reason to avoid abusing power, the strength of which will depend on the nature and structure of the practice in which the power is embedded.

Of course, much work remains to be done, as many of the concepts discussed here permit of further refinement. But I hope to have shown that the role abuse of power plays in popular conceptions of wrongdoing is not a chimera. Abuse of power is a genuine normative category—a category we have reason to avoid.[42]

*University of Kansas*
*ddorsey@ku.edu*

## References

Ceva, Emanuela, and Maria Paola Feretti. *Political Corruption: The Internal Enemy of Public Institutions*. Oxford University Press, 2021.

42 I would like to thank Brad Cokelet, Jason Raibley, and some very helpful anonymous reviewers for encouraging and helpful comments on this paper. I would also like to thank audiences at the University of Colorado, University of Pennsylvania, and University of Oxford for their kind indulgence of these ideas.

Dahl, Robert. "The Concept of Power." In *Political Power: A Reader in Theory and Research*, edited by Roderick Bell, David V. Edwards, and R. Harrison Wagner. Free Press, 1969.

Darwall, Stephen. *The Second-Person Standpoint*. Harvard University Press, 2006.

Fay, Brian. *Critical Social Science*. Cornell University Press, 1987.

Hart, H. L. A. *The Concept of Law*. Oxford University Press, 1961.

Jaggar, Alison, and Theresa Weynand Tobin. "Situating Moral Justification: Rethinking the Mission of Moral Epistemology." *Metaphilosophy* 44, no. 4 (2013): 383–408.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. Edited by Mary Gregor and Jens Timmerman. Cambridge University Press, 2012.

———. *The Metaphysics of Morals*. Edited by Lara Denis. Translated by Mary Gregor. Cambridge University Press, 2017.

Katz, Larissa. "Spite and Extortion: A Jurisdictional Principle of Abuse of Property Right." *Yale Law Journal* 122, no. 6 (2013): 1444–83.

Lewis, David. *Convention*. Blackwell, 1969.

Marmor, Andrei. *Social Conventions: From Language to Law*. Princeton University Press, 2009.

Miller, Seumas. "Corruption." In *Stanford Encyclopedia of Philosophy* (Winter 2018). https://plato.stanford.edu/archives/win2018/entries/corruption.

Morris, Herbert. "Persons and Punishment." *Monist* 52, no. 4 (1963): 475–501.

Nye, Joseph. "Corruption and Political Development: A Cost-Benefit Analysis." *American Political Science Review* 61, no. 2 (1967): 417–27.

Raday, Frances. "Privatising Human Rights and the Abuse of Power." *Canadian Journal of Law and Jurisprudence* 13, no. 1 (2000): 103–36.

Raz, Joseph. "The Rule of Law and Its Virtue." In *The Authority of Law: Essays on Law and Morality*. Oxford University Press, 1979.

Schroeder, Mark. *Slaves of the Passions*. Oxford University Press, 2007.

Stiegel, Lori. "Durable Power of Attorney Abuse: A National Center on Elder Abuse Fact Sheet for Consumers." American Bar Association, 2008. https://www.americanbar.org/content/dam/aba/administrative/law_aging/durable_poa_abuse_fact_sheet_consumers.pdf.

Vermeule, Adrian. "Optimal Abuse of Power." *Northwestern University Law Review* 109, no. 3 (2015): 673–94.

Vrendenburgh, Donald, and Yael Brender. "The Hierarchical Abuse of Power in Work Organizations." *Journal of Business Ethics* 17, no. 12 (1998): 1337–47.

Williams, Bernard. "Internal and External Reasons." In *Moral Luck: Philosophical Papers, 1973–1980*. Cambridge University Press, 1981.

*Journal of Ethics and Social Philosophy*
VOL. 30, NO. 2 · APRIL 2025
*https://doi.org/10.26556/jesp.v30i2.3519*
© 2025 Author

# CAN STATES RESIST MIGRATION BLACKMAIL WHILE PROTECTING MIGRANTS?

*Daniel Sharp*

States on Europe's periphery sometimes use the threat of creating a "migration crisis" to extract concessions from European Union (EU) member states.[1] This practice has become increasingly prevalent in recent decades due to the EU's increasing reliance on outsourcing migration control activities to neighboring states. To illustrate, consider the following cases:

- In 2010, Muammar Gaddafi, playing to racist fears among Europeans, threatened to "turn Europe black" with migrants from Africa unless the EU paid him five billion euros.[2] Gaddafi's regime in Libya was eventually deposed. However, the EU now pays hundreds of millions of euros to Libyan militias to pull back migrants, who are detained in dehumanizing prisons.[3]
- In 2021, Alexander Lukashenko, president of Belarus, sought to create the perception of a migration crisis along EU borders by encouraging migrants to transit through Belarus to Poland, Latvia, and Lithuania.[4] He, in conjunction with Russia, likely aimed to pressure the EU into repealing sanctions as well as to destabilize the EU in general. EU countries responded by suspending asylum procedures, increasing border security, engaging in mass pushbacks, and imposing additional sanctions.[5]
- After years of acting as Europe's gatekeeper under the 2016 EU-Türkiye Statement, Recep Tayyip Erdoğan announced in 2020 that Türkiye would no longer prevent migrants from crossing into Europe and would stop accepting migrants returned from the EU.[6] Erdoğan likely aimed to induce the EU to support its military actions in northern Syria and to recruit

1 In this paper, I use 'migrants' as a catchall category to mean people who (may) move across borders. This includes refugees, asylum seekers, and nationals of a blackmailing state if they are likely to move. I focus on the case of EU member states in this paper.

2 Adamson and Tsourapas, "Migration Diplomacy in World Politics," 123–24.

3 Hayden, *My Fourth Time, We Drowned*.

4 Halemba, "Europe in the Woods."

5 Ganty et al., "EU Lawlessness Law at the EU-Belarusian Border."

6 Muftuler-Bac, "Turkey and the European Union Refugee Deal."

additional funding (nominally, for hosting refugees) over which Türkiye had greater discretionary control. Greece responded by temporarily suspending asylum procedures, engaging in mass pushbacks, and reinforcing its borders (with EU support).[7]

- In 2011, the European Court of Human Rights ruled that EU member states should halt the return of asylum seekers to Greece due to human rights violations by Greek authorities.[8] Because deportation back to Greece was now more legally difficult, "Greece theoretically received a *carte blanche* to wave asylum seekers through its territory."[9] In 2015, the SYRIZA-ANEL coalition government adopted a policy of waving asylum seekers onwards, partly to gain leverage in its debt negotiations. Greece's defense minister, playing up racialized fears, declared, "we cannot keep ISIS out if the EU keeps bullying us."[10] Other EU member states responded by making agreements with Türkiye and various Balkan states. The European Commission then determined that asylum transfers could resume.[11] Its leverage undercut, Greece shifted its stance.
- The Kenyan government has repeatedly threatened the closure of the Dadaab refugee camp and the mass expulsion of refugees partly in order to recruit additional funding from international donors. Kenya has obtained around three hundred million dollars in aid from the United States, United Kingdom, and EU through this and other tactics.[12]

These are cases of noncooperative bargaining, wherein states leverage migration against target states who fear irregular migration. I call this phenomenon *migration blackmail*.

> *Migration blackmail*: State *A* makes a migration-related threat against state *B* in order to extract (unrelated) concessions from *B*.[13]

7 For an overview of Türkiye's migration diplomacy vis-à-vis the EU, see Kleist, "Beyond the Crisis Mode of the EU-Turkey Refugee Agreement"; Laube, "Diplomatic Side-Effects of the EU's Externalization of Border Control and the Emerging Role of 'Transit States' in Migration Diplomacy"; and Tsourapas, "The Syrian Refugee Crisis and Foreign Policy Decision-Making in Jordan, Lebanon, and Turkey."

8 European Court of Human Rights, *MSS v. Belgium and Greece* (Judgment), Application No. 30696/09, January 21, 2011, https://hudoc.echr.coe.int/fre?i=001-103050.

9 Tsourapas and Zartaloudis, "Leveraging the European Refugee Crisis," 250.

10 Tsourapas and Zartaloudis, "Leveraging the European Refugee Crisis," 245.

11 See Tsourapas and Zartaloudis, "Leveraging the European Refugee Crisis," 255–56.

12 Micinski, "Threats, Deportability and Aid," 2.

13 Migration blackmail is sometimes also discussed as "weaponizing migration." See Greenhill, *Weapons of Mass Migration*. I think this is a problematic frame. It wrongfully frames migrants as weapons. For a critique of the weaponization paradigm, see Bender, "Against

Migration blackmail comes in many varieties. However, in this paper, I focus on a "standard scenario": a blackmailing state (implicitly) threatens to flood a target state (or states) with unwanted migrants unless the target state (or states) meets certain demands. Specifically, the blackmailing state (typically a transit or host state) indicates its intention to allow a considerable number of migrants (typically a mixed migration flow) to transit onwards to the target's territory or external border unless the target (typically a regional neighbor or a union of states such as the EU) meets the blackmailing state's economic or political demands.[14] If the target state acquiesces, the blackmailer indicates it will slow migration by closing down the migration route; otherwise, the blackmailer indicates its willingness to manufacture a migration crisis at the borders of the target and often to place migrants in a position of acute vulnerability.

Migration blackmail may pose a dilemma for target states. This is because it seems to generate an especially acute conflict between two important goals that may appear to be, at first glance, mutually exclusive.[15] On the one hand, target states have interests in resisting and avoiding blackmail. This seems to require closing borders with the blackmailing state. Anything else means allowing the blackmailer to impose burdens on the target. On the other hand, migrants have significant interests, which are placed in jeopardy by this response. Many migrants have interests in securing access to international protection, and all have interests in avoiding acute vulnerability and harm. In cases where migrants will not receive adequate protection within the borders of the blackmailing state and lack reasonable alternatives to move elsewhere, the only way to

'Weaponised Migration'." I therefore prefer the term *migration blackmail*. I borrow this term from the migration diplomacy literature, which contrasts it with "backscratching," which is "promising to refrain from taking unilateral action against refugee populations within their borders, if compensated" (Tsourapas, "The Syrian Refugee Crisis and Foreign Policy Decision-Making in Jordan, Lebanon, and Turkey," 465). What distinguishes migration blackmail from mere noncooperation is the leveraging of migration to achieve some other non-migration-related goal. In defining migration blackmail to include only unrelated concessions (rather than migration-related concessions), I may depart slightly from Tsourapas's definition. This departure is relevant to the argument in section 5 below.

14 Not all cases of migration blackmail fit the standard scenario, as the case of Kenya illustrates. I focus on the standard scenario because it raises the dilemma most acutely.

15 Bauböck et al. define hard ethical dilemmas as follows: "Dilemmas are *ethical* ones if they involve choices between morally worthy goals that cannot be easily ranked, and they are *hard* if they cannot be easily resolved through clear thinking but persist in some form even after taking those actions that one considers as morally required or recommended" ("The Ethics of Migration Policy Dilemmas," 429–30). I follow their approach. Note that this definition of a dilemma is different from standard definitions of moral dilemmas, which require a conflict of obligations. Importantly, I argue below that not all cases of migration blackmail in fact give rise to hard ethical dilemmas.

respect their interests may seem to be to open borders in a limited way. It may thus seem that the target state must either give in to blackmail, thereby accepting the responsibilities the blackmailer imposes and their costs, or refuse those responsibilities, thereby leaving migrants unprotected or harmfully pushing them back across the border. None of these options seems wholly satisfactory. Thus, target states face the *migration blackmail dilemma*: there is no way to effectively resist blackmail while respecting migrants' legitimate interests in accessing protection and avoiding harm.

This paper examines the circumstances that can lead to this dilemma and considers how target states should navigate it. My core claim is that in a standard migration blackmail scenario, states are not justified in closing their borders or derogating from their protection obligations. This is an important conclusion because this has been the standard response taken by EU member states to migration blackmail—a response that has been legally sanctioned and enabled by the latest reforms to the EU Common European Asylum System.[16]

My argument proceeds as follows. I begin in section 1 by identifying what is problematic about migration blackmail. I then use this analysis to show in section 2 how migration blackmail may generate a dilemma. Importantly, these sections show that although migration blackmail may generate a dilemma for target states, it does not *always* do so. I then argue in section 3 that even when migration blackmail is dilemmatic, states often bear significant culpability for the situation. Still, even where migration blackmail is dilemmatic and states are not culpable, I argue that migrants' interests in accessing protection and avoiding harm trump states' interests in blackmail avoidance. This entails that closed-border reactions to blackmail are impermissible, and states are therefore required to adopt a stance of qualified openness (section 4). However, I contend that states retain a range of policy options to resist blackmail and to mitigate its costs even if they must honor migrants' claims to protection; permitting restrictive border closure is not necessary to resolve the dilemma (section 5). I briefly conclude by explaining the relevance of my argument for policy developments within the European Union.

## 1. WHAT IS WRONG WITH MIGRATION BLACKMAIL?

As defined above, migration blackmail occurs when one state uses the (potential) presence of migrants in its territory as leverage to extract unrelated

16 See European Commission, "Common European Asylum System," https://home-affairs.ec.europa.eu/policies/migration-and-asylum/common-european-asylum-system_en, (accessed February 10, 2025).

concessions from a target state or states.[17] Before considering how states are permitted to respond to blackmail, I first consider what may be morally problematic about migration blackmail as a practice. I do so because how states may respond to migration blackmail depends on an assessment of the permissibility of the practice.[18] Plausibly, migration blackmail is problematic when and because it either (1) wrongs migrants or (2) wrongs target states and their citizens. I consider each case in turn.[19]

One prominent concern is that migration blackmail *instrumentalizes* migrants and therefore wrongs them. Blackmail may do this because it involves issue linkage: using migrants as leverage for some external goal without regard for migrants' aims and interests. Plausibly, instrumentalization, which involves using migrants as a mere means rather than treating them also as ends in themselves, is problematic because it expresses the attitude that migrants are not entitled to the dignity due to agents.[20] This worry is valid. However, it is hardly unique to migration blackmail, and its moral weight is difficult to assess.[21]

Second, blackmailers often wrongfully threaten to harm migrants. Migration blackmail sometimes involves embedded threats, wherein the blackmailer *directly* threatens migrants in order to threaten target states. For example, Belarusian border guards reportedly beat migrants, force them across the Polish border, and do not allow them back into Belarus.[22] Analogously, Kenya would wrongfully harm Somali refugees if it expelled them; so, it is wrong to threaten to expel them. While the ultimate target of Kenya's threat to expel Somali

17 There is a fine line between migration blackmail and other forms of migration diplomacy, especially because almost all diplomacy issue linkage, and often the blackmailer's aims, may be opaque. What sets migration blackmail apart is the use of an implicit threat that leverages migrants under a negative frame from a migration-independent aim.

18 Below, I assume that legitimate states have a limited right to exclude because my analysis is largely addressed to policymakers in actual states operating under significant political constraints. Defenders of open borders may still share some of the concerns discussed about wronging migrants and target states. More importantly, they should endorse my core conclusions about the limits of how states may permissibly respond to migration blackmail.

19 Migration blackmail is most often employed by comparatively weak states. (Russia's recent actions vis-à-vis Finland are an exception to this rule, and Russian policy interests also play a role in the Poland-Belarus case.) However, more powerful states also employ problematic strategies (e.g., coercive threats and problematic conditionality) designed to induce weak states to accept unjust migration deals. Many of these strategies are subject to similar and often weightier objections. I discuss this issue in Sharp, "Collective Self-Determination and Externalized Border Control."

20 For discussion, see Mieth and Williams, "Beyond (Non)-Instrumentalization."

21 Scanlon, *Moral Dimensions*, ch. 3.

22 Perkowska, "(No) Children's Rights at the Polish-Belarusian Border," 251.

refugees is the international community, Kenya nevertheless wrongs these refugees by threatening to harm them. However, not all cases of migration blackmail involve threatening migrants. Türkiye's threat against the EU, for example, involved threatening to *stop* coercing migrants by allowing them to move onwards to Greece.

Such cases may be problematic for a third reason: namely, that they (deceptively) place migrants at serious *risk* of harm. For example, Belarus encouraged migrants to travel to Belarus with the promise that they would have access to the EU by crossing from Belarus to (e.g.) Poland, refused to offer these migrants protection in Belarus, and knew that it was unlikely that they would be admitted by EU member states. This foreseeably resulted in people freezing to death at the EU's external borders.[23] This has caused eighty-eight documented deaths since September 2021 at the Polish-Belarus border alone.[24] Belarus thus wrongfully subjected migrants to the risk of harm. Crucially, as I argue later, the ultimate harm in this case was coperpetrated by Poland, as these deaths arose only because Poland and other EU member states denied entry to migrants through violent pushbacks, physical border fortifications, and border closures.[25]

Migration blackmail therefore often—perhaps always—wrongs migrants.[26] One might therefore think that migrants have an interest in being protected against the practice. However, migrants also have interests in accessing protection, and, as we will see, these two goals may conflict.

Migration blackmail may also wrong target states. First, it might impose unfair burdens on the target state. Second, it might constitute a wrongful threat.

23 Kazharski, "An Authoritarian Spectacle."

24 See We Are Monitoring Association (Border Group Coalition), "Interactive Dashboards" (updated weekly), https://wearemonitoring.org.pl/en/statistics/interactive-dashboards/?cn-reloaded=1, accessed February 15, 2025.

25 Ganty et al., "EU Lawlessness Law at the EU-Belarusian Border."

26 It is an open question whether migration blackmail always *wrongfully* instrumentalizes migrants. Some migrants—refugees and those in similar positions—have interests in accessing international protection. If blackmail successfully facilitates entry to a target state willing to offer protection, then migrants may have a countervailing interest in allowing it to continue. Of course, migrants should not have to incur such risks to access protection. However, in the absence of routes to protection not based on blackmail, overburdened host states may be permitted to put pressure on target states by allowing or facilitating migration *so that they can better support migrants and refuges* within their territories. This arguably does not count as migration blackmail, strictly speaking, since such states do not try to extract *additional* migration-external benefits from those states by using migrants, beyond those benefits to the migrants themselves. Thus, although there are cases in which leveraging migrants in order to win financial support to host them may be permissible, migration blackmail is still likely wrongful in such cases because it involves using migrants for additional purposes.

Again, these worries do not apply to all cases of blackmail. The first assumes that protecting migrants comes with net burdens. Migration sometimes creates burdens, such as "objective" (e.g., financial) burdens, "subjective" burdens (e.g., unwanted changes to the target state's social landscape), the normative burdens of discharging obligations towards migrants, and interactive burdens that stem primarily from hostile reactions of the target state's population towards migrants (e.g., backlash).[27] The significance of these costs varies, but migration also comes with benefits, or at least opportunities for host states to benefit.[28] These potential benefits need to be taken into account in assessing the charge of unfairness, as threatening to benefit someone is not wrongful.

Both these wrongs presuppose that the blackmailer's threat imposes an *unfair* distribution of burdens. A threat to flood the target state with migrants wrongs that state when it leaves the state's choice situation worse than what it is entitled to from the blackmailing state.[29] Whether this is so depends on the background distribution of the burdens of migration. Importantly, target states usually do less than their fair share of refugee protection, while some blackmailing states, such as Türkiye, may do more than their fair share. Granted, a target state may have an interest in choosing *how* to discharge their migration-related duties. However, since it is often clear that a target state has no plans to discharge its duties at all, it is difficult to assign much weight to this interest. In most cases, this makes the appeal to fairness as a wrong-making feature seem hypocritical and empty, although there may be some cases in which the complaint may be reasonable. Poland, for example, hosts a large number of Ukrainian refugees, while Belarus hosts none (and is indeed partially responsible for their displacement).[30]

Moreover, migration blackmail is problematic when the blackmailer unjustly constrains the target states' *effective sovereignty*—the state's control over its internal socioeconomic dynamics and its reasonable freedom from external interference.[31] This may infringe the target state's self-determination. A blackmailing state may trigger these costs by threatening to overload the target state's capacity or to trigger political instability. Effective sovereignty

27 Kapelner, "Anti-Immigrant Backlash."

28 Betts, *The Wealth of Refugees*.

29 Kolodny, "What Makes Threats Wrong?"

30 For criticism of Poland's racialized refugee protection regime, which has been (relatively albeit imperfectly) welcoming to white Ukrainians while unleashing violent exclusion against migrants entering via Belarus who are racialized as nonwhite, see Balogun, "Refugees Separated by the Global Color Line."

31 Ronzoni, "The Global Order."

has both an objective and a relational component.[32] This latter component is what makes the migration blackmail dilemma distinct from the more general potential conflict between states' and migrants' interests.

To expand, the relational concern about blackmail is about an *unequal* power relationship in which the blackmailer can impose burdens on the target at will, thereby undermining the target's effective sovereignty and constraining their self-determination. This is worrying if, by dominating the target state, the blackmailer dominates its inhabitants.[33] This worry is most salient when (1) the blackmailer possesses asymmetric power, and (2) the threat of large-scale migration would genuinely impair the target's capacity for self-rule.

However, these conditions are not always—indeed, not often—met. On the one hand, states may be justified in proportionately pressuring those who do not comply with their duties to do their fair share in refugee protection schemes. While this may set back the target state's interest in self-government, it does not *wrongfully* do so if states are (as I argue later) obligated to partake in such schemes. On the other hand, target states are often already in a position of superior power vis-à-vis blackmailing states and routinely exercise this power over them to ensure that *they* must bear the brunt of the burdens of refugee protection. When this is so, migration blackmail may thus be viewed as a potential exercise of *counterpower*. Specifically, migration blackmail may be viewed as a form of resistance to nonvoluntary or coerced forms of responsibility sharing, which *de facto* impose greater obligations on states in the Global South.[34] Finally, everyone has an interest in norms that promote justice-conducive cooperation among states. If blackmail undermines such norms, we all have reason to object to it. However, where such norms are already conspicuously absent, the complaint against flouting them has limited force.

Thus, migration blackmail may potentially wrong target states in several ways. Crucially, however, it does not always wrong target states. When it does depends on the details of the case. My own tentative assessment is as follows. On the one hand, all cases of migration blackmail likely wrong migrants

32 Importantly, these two conditions may come apart: Russia has engaged in migration blackmail vis-à-vis Finland; it is vastly more powerful than Finland, but the low number of migrants who have been involved in this case means that this poses no serious threat to Finland's state capacity.

33 Pettit, "A Republican Law of Peoples."

34 Assessing whether migration blackmail counts as counterpower or as dominating raises a complex issue. Türkiye, in its migration blackmail, targets the EU, which is a more powerful confederation. However, it uses this leverage directly against Greece, which it has greater power than. Plausibly, this counts as a use of counterpower vis-à-vis the European Union but a case of dominating power vis-à-vis Greece. I discuss this issue further in Sharp, "Collective Self-Determination and Externalized Border Control."

because they instrumentalize them, although not all cases involve the other additional harms to migrants identified above. On the other hand, many but not all cases of migration blackmail wrong target states and their citizens. Some cases of migration blackmail may be exercises of counterpower against states or groups of states (e.g., the European Union) that pose no serious threat to the target state's capacity, while other cases do indeed risk wrongfully dominating or foisting unfair burdens on the target.

Whether blackmail wrongs the target state matters because it affects how states are permitted to respond to the practice. A state has a distinctive and strong reason to resist migration blackmail in ways that may set back migrants' interests *only if* the state in fact has legitimate and strong interests in avoiding blackmail in the particular case. To the extent that this is not the case, the dilemma I explore below does not arise. There may, however, be reasonable disagreement about these matters. My aim below is to address *reasonable* policymakers in target states who may feel the pressure that migration blackmail generates but who aim to respond with a sensitivity to the precarity of the situation of migrants. I thus assume below, unless otherwise stated, that migration blackmail wrongs both states and migrants. I assume this because it represents a charitable reconstruction of the reasonable moral concerns that policymakers in target states might have about migration blackmail. If it turns out that the above objections to blackmail are *not* applicable in a given case, this makes the argument for my normative conclusions more straightforward because it means the state interests that closed-border strategies seek to protect are less severely imperiled. Thus, my aim is to discuss and refute the best-case scenario for defenders of the harsh responses to migration blackmail that are sometimes advocated in policy discussions and to take (the reasonable bits of) policymakers' perspectives seriously. I discuss the nature of the dilemma that migration blackmail gives rise to in more detail in the next section.

## 2. The Migration Blackmail Dilemma

Wrongful migration blackmail may sometimes generate an apparent dilemma for target states. States have significant interests in avoiding wrongful threats and domination, avoiding unfair burdens, and maintaining justice-conducive cooperative norms. This gives target states reason to resist and avoid blackmail. A blackmailing state aims to demonstrate that it has the power to impose burdens on its target. Thus, it is important not only that states resist blackmail but that target states avoid blackmail by making themselves invulnerable to it. This latter objective seems to require that target states refuse the blackmailer's demands *and* avoid the costs that the blackmailer seeks to impose. States'

interests may thus appear to be best protected by adopting a closed-border response because even when the target state refuses the blackmailer's demands and accepts the burdens of migration, the target remains vulnerable to the blackmailer's ability to impose costs on them at will.

Yet migrants also have significant interests. Chief among these are interests in securing access to adequate protection and avoiding harm. Migrants are usually unlikely to receive adequate protection in the blackmailing state, as many of these states offer low-quality protection or no protection at all. Moreover, a closed-border strategy typically involves harmful and violent methods for keeping migrants from reaching the state's borders, such as mass pushbacks and pullbacks. These methods directly harm migrants.[35] For example, in response to Turkish blackmail, Greece hardened its external borders and engaged in violent pushbacks.[36] Similar events occurred at the Polish border.[37] As a result of these dynamics, migrants, regardless of status, are thereby placed in a position of extreme vulnerability. They are, practically speaking, stranded in the blackmailing state without protection. Many migrants literally cannot return to their states of origin via different routes. (They may lack the means to do so, and sometimes the blackmailing state prevents them from doing so.) Migrants with claims to refugee status also face persecution or, on broader definitions of who counts as a refugee, conditions under which their basic rights would be at risk of violation if they return to their states of origin. Thus, simply refusing blackmail—which in practice means either pushing back migrants or swiftly deporting them to their home states—fails to respect migrants' significant interests.

Therefore, respecting migrants' interests seems to require that a target state allow individuals being leveraged by blackmailers to enter and seek protection in its territory because these migrants usually have no reasonable means of accessing protection other than by entering the target state. In addition to these especially weighty interests, migrants have additional interests in exercising control over their lives—in controlling their own mobility and in receiving accommodation in a state that fits their preferences or interests—which further bolsters this claim.[38] Yet embracing limited openness means accepting that

35 Hillier-Smith, "Doing and Allowing Harm to Refugees" and *The Ethics of State Responses to Refugees*; and Schmid, "Saving Migrants' Basic Human Rights from Sovereign Rule."

36 Koros, "The Normalization of Pushbacks in Greece."

37 Grześkowiak, "The 'Guardian of the Treaties' Is No More?"

38 Although I believe this interest in autonomy is of crucial importance, I bracket it for purposes of my discussion, as it is likely to be contested by policymakers. Once one considers these and other interests in mobility, the case against closed-border responses becomes even stronger.

the blackmailing state can impose obligations on the target, which is less than what fully resisting blackmail seems to require. Thus, we arrive at the migration blackmail dilemma.

> *Migration blackmail dilemma*: There is no way for target states to effectively resist and avoid blackmail while adequately respecting migrants' interests in accessing protection and avoiding harm.[39]

It is unfair to insist that migrants' claims to international protection must go unmet to shield states against blackmail; however, requiring target states to open their borders or make concessions simply because another state decides to instrumentalize migrants seems to license a problematic norm of interstate relations and foist undue burdens on target states.

How, then, are states morally permitted to respond to blackmail? One must distinguish between two kinds of responses. States can respond proactively—by adopting *preventative solutions* that seek to stop blackmail from arising and by undercutting the circumstances that give rise to it; or they can respond reactively—by grappling with blackmail once it has arisen. While I will focus primarily on the latter issue in this paper, a discussion of the latter kind of response without the former would be shortsighted. First, it is desirable from the perspective of target states that they immunize themselves against blackmail. Second, purely reactive migration policies risk perpetuating blackmail by contributing to the dynamics that cause it.[40] Third, considering only reactive responses risks naturalizing the conditions that give rise to migration blackmail rather than seeing them as something that arises due to the policies of specific states. Finally, I argue in the next section that because the conditions that make blackmail possible or effective often arise due to the problematic policies of

39 I assume that many if not most of the migrants in question in blackmail scenarios have legitimate refugee claims. This is empirically plausible on any reasonable conception of refugeehood, but it is particularly plausible on broader conceptions of who counts as a refugee. For discussion of refugee definitions, see Shacknove, "Who Is a Refugee?"; Lister, "Who Are Refugees?"; Bender, "What's Political about Political Refugeehood?"; Buxton, "The Duty to Naturalise Refugees"; and Owen, "Differentiating Refugees." Importantly, many migrants who do not qualify for refugee status on the narrower definition outlined in the Geneva Convention are still extremely vulnerable, especially in cases of migration blackmail. Moreover, it is difficult to determine who is entitled to refugee status without assessing their claims; hence, people have a right to *seek* asylum. For those who lack valid refugee claims, states can, I assume for purposes of argument, generally avoid the costs associated with protecting them by simply deporting them to their home states after fairly assessing their claims. So migrants without claims to refugee status impose only limited costs on target states.

40 Kleist, "Beyond the Crisis Mode of the EU-Turkey Refugee Agreement."

target states, target states may sometimes be partially liable for some of the burdens associated with migration blackmail.

## 3. THE CULPABILITY ARGUMENT

Before exploring how states should navigate the migration blackmail dilemma, it is worth considering why migration blackmail occurs in the first place. Certain features of the international order make blackmail an attractive strategy for peripheral states.[41] These include large-scale displacement crises for blackmailing states to leverage, the negative (and often deeply racist) perceptions of (certain) migrants, the absence of fair responsibility-sharing schemes, and the broader policy trend towards the externalization of migration control. These conditions and circumstances are not natural or inevitable; instead, they often arise partly due to unjust actions on the part of the international community of states in general and on the part of target states in particular. When this is the case, these states are partially liable for the consequences of migration blackmail. At the very least, they lack the standing to complain about being coerced into protecting migrants. I call this the *culpability argument* for target states' special responsibilities in cases of migration blackmail.

There are several broader dynamics that have greatly contributed to the current situation in which migration blackmail appears as an attractive strategy to states on Europe's periphery. A first precondition for migration blackmail is the existence of a large number of forcibly displaced or otherwise vulnerable people in the Global South. A "supply" of people in this position is usually necessary for migration blackmail to appear as a feasible strategy. While some forced displacement originates from crisis and circumstances for which the origin state bears sole culpability, other cases of forced displacement are partially driven by serious injustices *for which target states bear some culpability*. Forced displacement from Afghanistan and Iraq, for example, is something for which many European member states bear some culpability. More indirectly, European diplomatic politics and economic policy may play a role in shaping displacement elsewhere in the Global South.[42] This grounds special responsibilities on the part of states towards these displaced persons. When this is so, these states cannot reasonably complain about being made responsible for these persons by a blackmailing state. Indeed, more broadly, states have independent moral and

41 The deeper background conditions that make migration blackmail possible include global poverty, global inequality, state dysfunction, a state system in which actors have asymmetric power, and the unaddressed legacies of colonial domination. These conditions are, in my view, conditions of ongoing injustice that states also have collective obligations to address.

42 Souter, *Asylum as Reparation*.

prudential reasons to find durable solutions for currently displaced individuals and at least to stop contributing to displacement in the Global South because doing so would help to prevent migration blackmail from arising.

To be sure, not all cases of displacement are ones for which target states specifically bear culpability. However, there are also certain background conditions that make migration blackmail possible, and these states may bear some responsibility for those. These background conditions include the overall structure and character of the global or regional migration regime that make migration blackmail appear to peripheral states as a viable and attractive strategy. First, states are susceptible to blackmail because of the securitization of migration, xenophobia, and racialized nationalism. These dynamics make certain forms of migration, even irregular (but not unlawful) entry by asylum seekers, appear as threatening. This is what gives blackmailers leverage. These perceptions, however, are not immutable. Indeed, some states and their political leaders, such as Poland, Greece, and many other EU member states, have greatly contributed to exacerbating these negative perceptions. These states bear some culpability for fomenting anti-immigrant attitudes that make migration blackmail appear so threatening and therefore a rational strategy for blackmailing states. States therefore have reason to engage in reframing migration (e.g., avoiding securitization and creating alternative narratives) and fashioning more inclusive national identities. They should also build arrival systems that can efficiently and fairly process and integrate new arrivals. Such systems preempt blackmail by reducing crisis perceptions and allowing states to better harness migrants' human capital, turning perceived burdens into economic and social benefits.

Second, a structural cause of blackmail is the absence of a system of fair responsibility sharing that justly distributes the burdens of migration governance in general and of protecting refugees in particular. Under such a system, there would be a clear mechanism for ensuring that the burdens of migration blackmail do not fall only on one state. Such schemes might even be designed in ways that specifically provide for solidaristic resettlement in cases of migration blackmail. States have duties to build just responsibility-sharing systems because this is necessary to sustainably guarantee access to protection and ensure interstate justice.[43] Instead of building such a system, however, states who are targets of migration blackmail, such as Poland, have worked actively against it.[44] This may impact the duties of states in this position. Specifically,

43 This extends slightly the argument in Aleinikoff and Owen, "Refugee Protection," 470–71. The argument is institutional: if states have a duty to ensure refugees are protected, then they also have duties to create institutions that reliably ensure their protection needs are met.

44 Poland has rejected introducing a robust responsibility sharing mechanism into the EU's common asylum system. See Vaagland and Chmiel, "Parochialism and Non-Cooperation."

it undercuts the standing of states like Poland to complain when they end up having to bear undue burdens.

Finally, European states have collectively pursued a general strategy of migration control that has made migration blackmail all but inevitable. Specifically, EU member states have coordinated to adopt a system of migration control based on containment and externalization. States in the Global North have created an elaborate architecture of remote control, which prevents asylum seekers from reaching their territories.[45] This system *de facto* outsources protection for refugees to third countries, often in the Global South, and also relies on these countries to engage in border control activities on behalf of the EU. Not only has this system been deleterious for refugees' rights; but as a direct consequence of this strategy, many people in need of refuge are unable to access adequate protection and many states in the Global South have become unfairly burdened as hosts.[46]

Crucially, outsourcing protection has made the EU increasingly dependent on third countries for border control and has led to an increased concentration of migrants in those states. This enables blackmail and makes it an attractive tactic of migration diplomacy.[47] The EU-Türkiye pact, for example, increased Türkiye's leverage vis-à-vis the EU, and as a result of its position as a major refugee host on the EU's periphery, Türkiye periodically threatens to open its borders in order to gain political and economic benefits from the EU. States thus have reason to end externalization policies in order to undercut blackmailers' leverage. In addition, given that target states are sometimes culpable for creating this dangerous dynamic in the first place, it may be reasonable to expect them to bear some responsibility for hosting refugees (who, as a direct result of the externalization policies of those states, have often languished for years with precarious (non-)protection) and for the foreseeable consequences of designing such a system in the first place.

I have argued that target states are often responsible for the conditions that give rise to migration blackmail. This is because target states often have (1) contributed to forced displacement, (2) contributed to wrongful perceptions of migrants as threatening, (3) worked against fair burden-sharing schemes, and (4) contributed to migrants' lack of access to international protection. Thus, target states bear some responsibility for creating the circumstances that give

45 FitzGerald, *Refuge Beyond Reach*; and Sharp, "Collective Self-Determination and Externalized Border Control."

46 For discussions of inadequate protection and overburdening host states in the Global South, see Parekh, *No Refuge*; Alienikoff and Owen, "Refugee Protection," 471; and Sharp, "Collective Self-Determination and Externalized Border Control."

47 Huysmans, "The European Union and the Securitization of Migration"; and Tittel-Mosser, "Reversed Conditionality in EU External Migration Policy."

rise to migration blackmail in the first place and for the precarity of migrants who seek protection. They may therefore lack the standing to complain when they are made to bear additional costs in the form of migration blackmail.

More generally, understanding the conditions that give rise to migration blackmail can also help us to understand that there are a range of justice-conducive ways that states might prevent migration blackmail in the first place—policies that states have duties to adopt anyway. These include (1) ending their complicity in forced displacement, (2) cultivating more welcoming attitudes towards immigrants, (3) instituting fair burden-sharing schemes, and (4) ending reliance on border externalization. These policies are not only morally desirable and arguably morally required; they also are *prudentially* in states' interests because they would help prevent migration blackmail.

Importantly, I do not want to overstate the force of the culpability argument. The fact that target states bear some responsibility for the dynamics that give rise to migration blackmail does not morally permit blackmailers to blackmail. Blackmailing states bear significant culpability for their actions. Moreover, not all potential target states bear significant culpability for all cases of migration blackmail. For example, Finland may not bear significant culpability for Russian migration blackmail. Still, the culpability argument constitutes a helpful corrective to target states' portrayal of themselves as purely innocent bystanders who are targeted for no good reason by nefarious states beyond Europe's borders. Thus, in the next section, I offer an independent argument for why states should prioritize migrants' fundamental interests in responding to migration blackmail. This argument applies even where target states do not bear special responsibility for creating the conditions under which migration blackmail occurs and even if one rejects the culpability argument.

## 4. THE PRIORITY CLAIM

The migration blackmail dilemma, to recall, is that it appears states cannot effectively resist blackmail while adequately respecting migrants' interests. How are states morally permitted to navigate this dilemma? I focus below on what states are morally permitted to do, not on what would be morally best for them to do. I base my discussion solely on minimal principles rather than on a full theory of migration justice. This is because migration justice is a matter about which there is serious disagreement, and these more minimal assumptions are more likely to appeal to policymakers. In this section, I defend a claim about the weight of the interests at stake and then show that this rules out standard responses to migration blackmail, which I call *closed-border responses*. The specific claim I defend in this section is:

> *The priority claim*: Migrants' interests in accessing adequate protection and avoiding harm ought to take precedence over state interests in blackmail avoidance.

Two things are worth noting about this claim. First, the priority claim is not a general claim that migrants' interests ought to take precedence over the interests of states and their citizens. Rather, it picks out particular interests—namely, in accessing adequate protection and harm avoidance—and claims that these take precedence over those state interests plausibly threatened by blackmail.[48] Second, my aim is to defend the priority claim *in the specific case of migration blackmail*. This involves showing specifically that migrants' interests in harm avoidance and accessing international protection take precedence over the state interests that are plausibly threatened by migration blackmail. What makes the priority claim interesting is that states and their citizens are often taken to have (and if the argument of section 1 is correct, may indeed sometimes have) distinctive interests in avoiding blackmail that go beyond their generic interests in exercising control over migration. Thus, I offer a defense of the priority claim that takes these interests seriously.

The priority claim may be supported by two lines of reasoning: one that appeals to migrants' interests in avoiding harm, the other that appeals to their need for protection. I sketch both arguments below.

### *4.1. The Harm Argument*

A first argument for the priority claim stems from the widely held view that there is an asymmetry between doing and allowing harm. On many deontological moral theories, it is worse to perpetrate harm than to fail to aid. Harming involves causing a person to be worse-off by bringing about a state of affairs that significantly impacts their core interests or well-being.[49] Yet excluding migrants in situations of migration blackmail almost always involves harming them.

This is so for two reasons. First, states perpetrate exclusion through literal threats of force and violent assaults on migrants. In the case of Poland, migrants are sometimes beaten and physically pushed behind a border fence by the Polish border guards. This is an uncontroversial case of harm.[50] Second,

48 Migrants may have other weighty interests and claims (such as autonomy and equality) that may also usually take precedence over the interests of citizens, even if states have special duties to their citizens. However, I do not rely on these claims here. For a discussion of equality, see Sharp, "Relational Equality and Immigration" and "The Right to Emigrate." For a discussion of autonomy, see Oberman, "Is There a Human Right to Immigrate?"

49 Hillier-Smith, "Doing and Allowing Harm to Refugees," 301–2.

50 Other times, they are detained in appalling conditions for long periods, something that is further enabled by recent changes to the European asylum system. See Majcher, "Creeping

migrants instrumentalized by blackmailing states are often caught in a uniquely vulnerable position because they (1) are literally prevented from returning to their states of origin or cannot return safely and (2) face a systematic condition of nonprotection or underprotection in the blackmailing state. Consider the following statement by a refugee about the situation in Poland:

> We want to stay here and become Polish citizens, but they do not allow us to do so. When we go to Belarus, they beat us there, take our money, and send us back to Poland. The Belarusian police expel us, to Poland. These children cannot walk, they will all die on the road, in the forest. We have no food or water. We walk 40 km to Belarus, and there they catch us, beat us, and send us back to Poland. And so on and on between Poland and Belarus.[51]

Cases like this count as not merely allowing but perpetrating harm against migrants. When one agent intentionally and forcefully prevents another from escaping a harm that may befall her—by, say, blocking a door that would allow her to escape her torturer—he causes her to be harmed. In the present case, the harm in question *essentially depends* on being denied entry to the target state. It is the behavior of the border guards in Poland that causes migrants to be in a state of nonprotection and extreme vulnerability. Belarus is thus not solely culpable for the harm; Poland is too.

Thus, when a state forecloses migrants' only means of accessing asylum, it harms those migrants. Such harm requires a particularly high threshold for justification, which migration blackmail does not meet. The Polish case is extreme because migrants literally cannot return to their country of origin, since Belarus prevents them. But even where this is not the case, most people caught in situations of migration blackmail are refugees or similarly necessitous migrants: for them, returning would place them in situations of danger or extreme precarity.[52]

The harm argument helps to respond to a likely objection to the priority claim—namely, that states are entitled or even obligated to privilege the interests of their compatriots and so to exclude migrants in cases of blackmail, even if this means failing to benefit migrants. The problem with this argument is that the exclusion of people in this case involves not simply failing to benefit migrants but actively harming them. But neither a state's special responsibilities towards its citizens nor one's associative duties towards one's compatriots can

---

Crimmigration in CEAS Reform."

51 Quoted in Perkowska, "(No) Children's Rights at the Polish-Belarusian Border," 251.

52 Importantly, many of these harms occur whether or not a person has a valid claim to refugee status.

justify harming others.[53] A parent may not harm innocent children in order to serve his own child's interests, and states may not aggress against other states and harm foreign citizens in order to serve their own citizens' interests, even if parents have special duties towards their children and states have special duties towards their citizens.[54] Analogously, the fact that states have special obligations towards their citizens does not permit them to harm migrants.

This general position can be further supported by considering in more detail the significance of citizens' interests in blackmail avoidance. As explained above, states' interests in blackmail avoidance include interests in avoiding unfair burdens and domination and in maintaining effective sovereignty. Yet states' interests ultimately matter when and because they impact the lives of the individuals who states represent. However, the fact that a state must bear unfair burdens does not imply that any group of its citizens is asked to bear significant burdens. Certainly, individuals have interests in living in a representative state that possesses effective sovereignty. Such an arrangement protects them against external domination; and so when a state interferes with another state, it potentially exposes the citizens of that state to some degree of domination. Yet the degree of domination in realistic cases of migration blackmail is usually minimal. In the case of Finland and Russia, one of the more compelling cases of migration blackmail, the actual burdens that Finnish citizens are being asked to bear is extremely limited. It is simply not the case that hosting an additional few more thousand refugees will impair essential interests of citizens in self-determination or foist upon them unreasonable costs. Migrants, in contrast, have clear, fundamental interests at stake in accessing adequate protection and in having their basic rights protected. These interests are simply weightier than those that blackmail threatens to impair for the citizens of target states.[55]

### *4.2. The Protection Argument*

The above argument for the priority claim appeals to the premise that excluding migrants in cases of blackmail involves harming them. But let us suppose for the sake of argument this is not the case. Suppose instead that the exclusion of migrants in cases of blackmail merely results in harm to migrants but does not

53 Hidalgo, "Associative Duties and Immigration."

54 Hillier-Smith, *The Ethics of State Responses to Refugees*, 67.

55 Perhaps the most significant interest at stake is that of maintaining core democratic institutions over time. Insofar as migration blackmail threatens to impair these, it provides a serious reason for concern. See Kapelner, "Anti-Immigrant Backlash." However, there is little reason to believe core democratic institutions are seriously threatened by most cases of migration blackmail. Where they are threatened, they are threatened by reactionary forces *within the host state*, not by migrants, and so it is unfair to displace this burden onto migrants.

involve harming migrants. Still, migrants' interests in accessing protection in cases of blackmail trump the interests of target state citizens in avoiding blackmail.

There are at least two broad ways to develop this argument. The first appeals to the role of refugee protection in securing the legitimacy of the international order.[56] The state system is a dispersed system of governance that assigns individuals whose basic rights need protection to states charged with protecting them. A condition of this arrangement being minimally justifiable to those it governs is that everyone's rights are reliably protected.[57] Yet states often fail to protect human rights. This creates a duty on the part of the community of states to repair this departure from the basic conditions of the legitimacy of the international order and to ensure each person's rights are protected. The international refugee protection system is just such a legitimacy repair mechanism. Because ensuring the protection of fundamental rights is a basic condition for the justifiability of the international order, this duty takes precedence over issues of fairness in the distribution of burdens of protection.[58] It takes precedence, specifically in this case, for two reasons. First, excluding refugees in cases of blackmail may leave them in a state of nonprotection. In fact, it constitutes nonrefoulement because it returns them to a situation in which their basic rights are at risk. But nonrefoulement is the central and most stringent norm of the international refugee regime and so takes precedence over fairly distributing the burdens of refugee protection.[59] Second, the legitimate power of states to exercise control over migration, on the legitimacy repair view, depends on states discharging their duties towards refugees.[60] But in cases of migration blackmail, these duties are not discharged because refugees remain unprotected. So states cannot, on this view, legitimately exclude refugees in situations of migration blackmail.

A second argument appeals to a more ecumenical justification of states' duties to refugees.[61] This is the view that the moral foundation of states' duties

56 Owen, "In Loco Civitatis" and *What Do We Owe to Refugees?*; Buxton and Draper, "Refugees, Membership, and State System Legitimacy"; and Sharp, "Immigration and State System Legitimacy."

57 Owen, *What Do We Owe to Refugees?* For an alternative legitimacy criterion, see Buxton, "The Duty to Naturalise Refugees."

58 Owen, "Refugees, Fairness and Taking Up the Slack."

59 There is some reasonable debate about how to understand nonrefoulement. The decisive moral issue is, however, whether refugees are returned to a place where their basic rights are placed at risk.

60 Sharp, "Immigration and State System Legitimacy."

61 I think both arguments for duties to protect refugees are convincing. I thus believe that this duty has multiple grounds. However, many philosophers may endorse only the latter ground, and so I consider these two positions independently.

towards refugees is that states have a duty to rescue those in need of protection. The priority claim, as a claim about the weight of the moral interests at stake, is not seriously disputed even by those who endorse a rescue-based view of states' obligations towards refugees. For example, although David Miller is skeptical about the demandingness of duties to rescue, this is not because he takes the interests of refugees to have less moral weight but because he is skeptical about *how* duties can be derived from refugees' interests.[62] Miller agrees that people's interests in receiving protection are very weighty. Indeed, their basic rights and most fundamental interests are at stake. Thus, on no understanding of the migration blackmail scenario do citizens have similarly weighty interests.

The issue, then, is how these interests ground a duty to rescue, not whether they are weightier than those of citizens. So, let us consider the two most widely asserted limits to the duty to rescue.[63] One alleged source of limits to duties to rescue is the collective nature of these duties. Specifically, the claim is that it might seem unfair to saddle *only* target states with such duties. But this argument is unconvincing for four reasons. First, the duty to effectively remedy injustice plausibly takes precedence over fairness in discharging this obligation among the duty holders, even in standard rescue scenarios. Second, as I have argued above, states that are targets of migration blackmail are often responsible for undermining efforts at fair sharing. Appeals to fairness thus seem uniquely inapplicable in these cases. Third, in many cases of blackmail, there are no alternative ways in which migrants can access protection other than by entering the territory of the destination state. When this is the case, the collective duty is *de facto* particularized. In practice, *only* the target state can engage in rescue. However, in the next section, I further argue that the costs of rescue may be distributed in a second stage, once migrants are granted a safe route out of their precarious situation.

Finally and most importantly, the appeal to fairness in fact explains why it is *more* problematic to exclude migrants in cases of migration blackmail. If the complaint here is about fairness, that same complaint applies even more starkly to the way closed-border responses treat migrants. Closed-border responses involve displacing the costs of avoiding blackmail onto migrants themselves.[64] But migrants are not culpable for their predicament. They are merely attempting

62 Miller, "The Nature and Limits of the Duty to Rescue" and "Responsibility and the Duty of Rescue."

63 For discussion, see Herrmann, "Cosmopolitanism, Global Justice and Refugees," 180–85. A further qualification of the duty to rescue that Herrmann discusses, the indeterminacy of the duty bearer, is less salient in this situation given that only the target state(s) can typically initially rescue individuals from vulnerability in cases of blackmail.

64 Compare Owen, "Refugees, Fairness and Taking up the Slack," 154.

to access international protection through one of the limited available means, and seeking asylum is neither a crime nor something for which migrants may be reasonably blamed. However, target states are part of the group that holds duties to rescue and among whom these duties should be appropriated. They are thus liable for taking up the slack when others in the community of duty bearers fail to do their part. Thus, displacing the costs onto the *victims* of instrumentalization—victims who have often already been made to bear the unfair burdens of a dysfunctional global protection regime—is especially unfair. It is substantively more unfair than asking that a state take on more than its fair share of the burdens of refugee protection. Considerations of fairness, then, speak decisively against closed-border responses.

A second qualification of the duty to rescue concerns costs. It might be thought that the costs that states are made to bear in some cases of migration blackmail are too high to reasonably hold them under a duty to protect migrants. There is, to be fair, a genuine moral dispute to be had about the costs of rescue. However, there are several reasons to believe that reasonable parties to this dispute should maintain that target states are under a duty even in cases of migration blackmail. First, states sometimes bear special responsibilities for protecting particular migrants in these cases. Second, real-world cases of migration blackmail do not pose an acute threat to core state functions or to the fundamental interests of citizens of target states. Third, even if a given target state must bear a disproportionate burden, this does not automatically entail that any individual citizen of that state must bear such a burden. Since state interests do not matter fundamentally, showing that individual citizens are unduly impacted is necessary for the argument to succeed. Finally, if the target state in question can substantially mitigate the costs of rescue, this would suffice to defeat worries about demandingness based on costs. In section 5, I argue that this is often the case, which helps to complete my defense of the priority claim.

### *4.3. The Priority Claim in Practice*

Suppose, then, that the priority claim holds. Some important implications for how states may respond to migration blackmail follow. A target state's response to migration blackmail divides along two dimensions: whether the target state *allows or refuses* migrants' entry and whether it *acquiesces to or resists* the blackmailer's demands. The priority claim rules out reactive strategies that involve denying migrants entry, at least in realistic scenarios of migration blackmail. These closed-border responses come in several varieties, and the priority claim generates a strong presumption against all of them.

*Closed-border defiance* occurs when a target state rejects the blackmailer's demand *and* denies migrants entry without securing protection for them

elsewhere. This has been one of the EU's standard responses to migration blackmail. Greece adopted such a response to Türkiye in 2020, and Poland adopted such a response to Belarus in 2021 and onwards. This approach is usually defended on grounds that it sends a clear signal that target states will not allow themselves to be subject to blackmail pressure. However, this approach is unjustifiable if, as is often the case, migrants' rights are not adequately protected in the blackmailing state. This violates the priority claim, since in scenarios of migration blackmail, migrants typically have no realistic option of receiving international protection other than by accessing the territory of the host state.[65] Second, closed-border defiance requires harming migrants. To ensure that migrants do not enter their territory, states engage in violent pushbacks that block escape routes to safety.[66] This displaces the costs of avoiding blackmail onto migrants, who are not culpable for the extortive behavior of blackmailing states. Finally, this response unfairly disadvantages migrants by preemptively rejecting their claims—indeed, thwarting their ability to lodge them—without assessment of their merits, as made most patent by Poland's recent suspension of the right to asylum, which has been *de facto* sanctioned by the EU.[67] Closed-border defiance thus resists blackmail at the expense of protection. This is not a permissible way to balance conflicting interests in cases of blackmail.

A second kind of closed-border response is *closed-border externalization*, wherein a target state complies with (the core of) the blackmailer's demands. In exchange, the blackmailer may agree to arranging externalized protection in its territory, although such agreements invariably involve other (e.g., financial and political) concessions as well, given that blackmailers also have migration-independent aims that motivate them to engage in blackmail in the first place. This strategy may allow for the target state to avoid some of the burdens that the blackmailer seeks to impose. Yet it may prove counterproductive. In negotiating externalized protection, the target state may cement its dependence on the blackmailer, leading to more blackmail in the long term. Moreover, meeting the blackmailers' demands may be impermissible. While it would have been permissible to cut Greece a favorable deal in debt negotiations, it would not have been permissible to support Türkiye's incursions in northern Syria. More

65 For example, it is not an option for migrants in Belarus to return (Belarus does not allow it) or to receive protection in Belarus (Belarus does not offer it). I say typically because there may be potential exceptions. However, even in cases that are sometimes pointed to as exceptions, such as that of Türkiye or Greece, the "protection" offered is limited, the conditions migrants face are dire, and many core rights go unprotected. Kleist, "Beyond the Crisis Mode of the EU-Turkey Refugee Agreement."

66 Border Violence Monitoring Network, *The Black Book of Pushbacks*.

67 Vinocur et al., "Poland Wins After EU Backs Its Proposed Asylum Ban for Russia, Belarus."

generally, target states should be extremely cautious about externalization agreements that bolster illegitimate actors in blackmailing states by providing them with funding and dual-use capacity.[68]

Closed-border externalization is thus potentially permissible only if the arrangements lead to adequate protection for migrants. However, as I have argued above, this is not typically the case. Whether externalization leads to adequate protection depends in part upon what kind of protection persons require.[69] Yet blackmailing states are typically poorly positioned to provide adequate protection. Some blackmailing states are unable to offer adequate protection because they are oppressive regimes. If political oppression is at least one of the conditions that grounds claims to refugeehood, then oppressive states cannot in principle fulfill the normative function of refugee protection.[70] Other blackmailing states simply lack the capacity or willingness to provide adequate protection. Even if their capacity could be augmented by international support, this is often feasible only in the medium term. Thus, externalization in practice is rife with protection gaps and depends on dehumanizing measures such as encampment and detention.[71] These measures fail to adequately respect migrants' fundamental interests. For this reason, closed-border externalization is, in practice, not permissible as a response to blackmail.[72]

A final ultimately unjustified response that is ruled out by the priority claim is the approach taken by the European Union under the instrumentalization provisions of the "Crisis Regulation," a part of the recent EU pact on migration and asylum that the EU formally adopted in April 2024.[73] This regulation essentially allows EU member states to derogate from their protection obligations and to assess asylum applications via so-called border procedures—rapid consideration of asylum claims in border areas with few legal protections and little

68 Jakob and Schlindwein, *Dictators as Gatekeepers for Europe*; and Sharp, "Collective Self-Determination and Externalized Border Control."

69 For further discussion, see Owen, *What Do We Owe to Refugees?* chs. 2–3; and Alienikoff and Owen, ""Refugee Protection," 470–72.

70 Bender, "What's Political About Refugeehood?"

71 Parekh, *No Refuge*; and Hillier-Smith, "Doing and Allowing Harm to Refugees."

72 This conclusion leaves open whether there may be cases in which externalized protection in the blackmailing state *in theory* might be permissible as a response to blackmail. Although I am skeptical about this, ruling this out is not necessary for my argument. In any case, states have strong prudential interests in not relying on this strategy—namely, it makes them vulnerable to migration blackmail in the future.

73 See Regulation (EU) 2024/1359 of the European Parliament and of the Council of 14 May 2024 addressing situations of crisis and force majeure in the field of migration and asylum and amending Regulation (EU) 2021/1147. For a detailed critique of a prior draft, see European Council on Refugees and Exiles (ECRE), "ECRE Reaction."

recourse for appeal coupled with the legal fiction of nonentry to limit migrants' rights and facilitate quick deportation.[74] It also dramatically expands the use of detention—often in inhumane conditions—for asylum seekers. These measures, it has widely been argued, are likely to result in refoulement, as they *de facto* sanction the expulsion of asylum seekers before they are given genuine hearings. While this approach is perhaps nominally more moderate than the straightforward closed-border strategies taken unilaterally (albeit with tacit EU support) by states like Poland, it nevertheless fails to take seriously asylum seekers' claims to apply for asylum and to access protection, despite the fact that they are within the territory of the target state. This approach again displaces costs onto migrants for few benefits, as detention and border procedures do not help to reasonably manage costs but instead counterproductively contribute to crisis perceptions and funnel people into costly and inhumane detention facilities.[75] Thus, the rhetorical appeal to instrumentalization—a concept vaguely defined in EU law—is used to justify punishing rather than protecting migrants who are the victims of instrumentalization and to undermine core legal protections for asylum seekers.[76] Since there is no further purpose that this category serves under current EU law, there is no compelling reason to codify this category in the EU's asylum regulations.

## 5. MITIGATING THE COSTS OF BLACKMAIL WHILE PROTECTING MIGRANTS

The claim that closed-border responses are impermissible leads to the provisional conclusion that in cases of migration blackmail, target states are not permitted to exclude and instead must allow migrants to access protection in their territories. Given that the question of whether migrants have such claims can typically be assessed only after allowing entry, target states must allow people to enter and lodge asylum claims.[77] This involves providing safe pathways

74 Grześkowiak, "The 'Guardian of the Treaties' Is No More?"

75 Majcher, "Creeping Crimmigration in CEAS Reform."

76 See Regulation (EU) 2024/1359 of the European Parliament and of the Council of 14 May 2024 addressing situations of crisis and force majeure in the field of migration and asylum and amending Regulation (EU) 2021/1147, especially section 4(b) for the definition of 'instrumentalisation'. For critical discussions, see ECRE, "ECRE Reaction"; Ganty et al., "EU Lawlessness Law at the EU-Belarusian Border"; and Ancite-Jepifánova, "Migrant Instrumentalisation."

77 What this means in practice depends on the specifics of the scenario. Minimally, it might involve the following: a target state might allow migrants currently residing in the blackmailing state to enter and lodge asylum applications; the applicants should then be allowed to remain in the state's territory (or the territory of a country where their rights will be adequately protected) while their claims are fairly assessed by a competent and unbiased

for entry from the blackmailing states for migrants with claims to protection, allowing them to formally apply for asylum, and fairly assessing their claims. This may seem to severely constrain the ability of target states to resist blackmail. Therefore, some may doubt whether states are indeed required to maintain such openness. I consider two ways of pressing these doubts below. In so doing, I demonstrate that target states have options for resisting blackmail that do not depend on displacing the costs of doing so onto vulnerable migrants.

### *5.1. Incentivizing Blackmail*

One might argue that my analysis underestimates the moral importance of resisting blackmail. I framed the migration blackmail dilemma as a conflict between the interests of states and the interests of migrants. However, one might argue that migrants also have interests in avoiding being instrumentalized. If limited openness incentivizes future blackmail, this may be net worse for (potential) migrants in the long run, and states may have reasons to take a harder stance in order to disincentivize instrumentalization.

This objection is analogous to a familiar objection to paying ransoms—namely, that it incentivizes ransom-taking.[78] A similar line of reasoning was one of the putative justifications for the EU's so-called Instrumentalization Regulation (a component of the larger Crisis Regulation)—namely, that a harsh response is needed to disincentivize instrumentalization.[79] Yet the objection is unfounded, and the parallel is inapt. Most fundamentally, punishing the victims of instrumentalization is not an appropriate response to concern about people being instrumentalized. But this is what the Instrumentalization Regulation in effect does.

Moreover, the argument turns on an implausible claim about incentives. I have not argued that states should capitulate to blackmailers. *That* would incentivize blackmail. Rather, I have argued that states should open their borders in a minimal way to allow vulnerable migrants to escape instrumentalization. Opening safe routes for migrants fleeing a blackmailing state does not incentivize blackmail.[80] Indeed, it does the opposite. Opening a safe migration route from the blackmailing state signals that the target state is impervious to black-

authority; if a claim is successful, the applicant should receive asylum, either within the state's borders or elsewhere as part of a just and fair responsibility sharing scheme.

78 Howard, "Kidnapped."

79 See Regulation (EU) 2024/1359 of the European Parliament and of the Council of 14 May 2024 addressing situations of crisis and force majeure in the field of migration and asylum and amending Regulation (EU) 2021/1147.

80 The strongest argument against paying ransoms is that doing so makes the payer of the ransom complicit in injustice (Howard, "Kidnapped"). This argument may apply to paying

mail. It signals that the target state is willing to accept the cost imposed without acquiescing to the blackmailer's demands. Third, opening safe routes that bypass a blackmailing state undermines the leverage that blackmailers need in the first place. This is an important preventative measure that target states might further take to insulate themselves from blackmail. Finally, opening safe routes is compatible with engaging in counter-blackmail measures, such as sanctions, to further deter blackmail. Because target states can impose counter costs on blackmailing states without imposing these costs on migrants, they can resist blackmail without displacing the costs of doing so onto migrants. I discuss this further below.

*5.2. Unfair Burdens Revisited*

A second objection to my argument is that limited openness unduly constrains the ability of target states to resist blackmail because it requires them to bear the burdens associated with migration. This might seem unfair. Must target states really accept that blackmailing states can impose obligations upon them at will? There are really three issues behind this objection, which I will treat in turn.

The first is whether the fact that an obligation is unjustly imposed invalidates that obligation. I believe not. Whether one agent can impose an obligation on another is distinct from whether the agent on whom that obligation is imposed has a complaint against its imposer. Consider an analogy. If Jim starts throwing babies in the pool, and Tina is the only capable swimmer around, Jim thereby triggers Tina's duty to rescue the babies. Tina has a complaint against Jim for compelling her to engage in rescue activities, but this complaint does not invalidate her obligation. Analogously, target states may have a complaint against blackmailers, but this does not invalidate their duty to protect vulnerable migrants. Why must target states bear the burdens of these rescue activities? The answer in the above case is that only Tina can rescue the babies from the pond, and this is the answer in the case of migration blackmail as well. There is usually no way to rescue those trapped in blackmail scenarios without letting them enter the target state's territory. If this means the target state must take on the burdens of protection activities, then so they must. This conclusion again is fully compatible with the target state having a fairness complaint against the blackmailer.

The second issue is whether target states lack a reasonable way to resist migration blackmail while maintaining limited openness for migrants from a blackmailing state who seek to enter their territories. Target states have other options to resist blackmail. On the one hand, target states can counter blackmail

a blackmailer in the context of an externalization agreement, but it does not apply to accepting migrants.

by other means. Specifically, they can do so by imposing costs on the blackmailing states in other domains. The standard way to do this is via diplomatic sanctions aimed at regime officials or via sanctions against states or individuals.[81] Such sanctions can plausibly disincentivize blackmail without excluding migrants. Such sanctions have an advantage: because other states in the target state's region typically have an interest in avoiding blackmail as well, it may be comparatively easier to form coalitions to sanction blackmailing states than it to organize burden-sharing schemes. Of course, sanctions presuppose that the target state or states have leverage over the blackmailer. Sanctions, particularly economic sanctions, may also prove ineffective and impose costs on civilians. However, sanctions can usually be targeted (e.g., towards elites) in ways that avoid these worries, and they are therefore morally preferable to resisting blackmail by inflicting harm on migrants. Moreover, as I have suggested above, it may be possible to signal resistance to blackmail by *opening* borders. Specifically, if the target state can credibly indicate that it does not regard accepting migration as a serious cost, this can disincentivize blackmail.

The third issue is whether limited openness entails accepting all the burdens associated with migration. Crucially, it need not entail this. This is because there are often strategies available to target states to decrease the burdens of allowing entry and offering protection. First, migrants can benefit host states when protection is organized effectively.[82] At least, states can organize protection in ways that minimize the costs of providing it. They can adopt policies of social and economic integration and offer migrants training in important sectors of the economy where labor is in demand. Second, by seeking to reframe migration positively, they can try to decrease xenophobic public reactions, which are a main source of the costs in these cases. These mitigation strategies may prove difficult to implement due to hostile political climates in target states. However, when such options are open to target states, it is simply not the case that accepting migration imposes only burdens on host societies, and to the extent that it does, this is partially attributable to social and political failures in those societies.

Moreover, maintaining a stance of limited openness is not the same as offering all who enter protection. The basic duty of states in circumstances of migration blackmail is to *facilitate* access to protection. A target state may do this by

81 My aim here is not to defend any particular policy but merely to indicate the existence of alternative means of resisting blackmail. Asset freezing is an example of a standard sanction against individuals (e.g., regime officials); economic and diplomatic sanctions are examples of measures deployed against governments.

82 Betts, *The Wealth of Refugees*; and Gowayed, *Refuge*. This is but an instance of a wider truth that immigrants generally benefit host states economically.

offering protection to those in need, but it might in principle do so in other ways. Specifically, the state might rely on burden-sharing agreements, either with other target states or through regional schemes. Doing so would allow the target state to diffuse the blackmailer's leverage *and* to avoid bearing the full cost of protecting migrants. I have suggested above that states are independently required to institute fair responsibility-sharing schemes.[83] However, because such schemes have proven difficult to realize, the key question is how target states might facilitate access to protection while shirking unfair burdens in their absence.

States may do so in cases of migration blackmail through a mix of positive and negative bargaining strategies. Positively, states might appeal for solidarity. States may thus seek to recruit additional funding to shoulder the burdens of hosting migrants. This type of bargaining strategy is sometimes called "migration backscratching."[84] This involves a state signaling a willingness to host migrants if they receive adequate support for doing so. This strategy can be effective in recruiting funds. It may be particularly likely to succeed in the EU context. EU member states have strong desires to reduce regional migration pressures, and few EU states indicate that they are willing hosts. Other EU member states may thus be more willing to fund protection to prevent future blackmail if a target state indicates its willingness to act as a host for a sizable portion of the migrant population in question. States might also seek to encourage other states to offer resettlement. They can do so by signaling that they are willing to do their part and to share in the burdens of refugee protection.

Negatively, target states may engage in limited noncooperation. Less controversially, states in the EU may simply allow some number of migrants to transit onward, in effect displacing costs of protection onto a neighbor. This is what Greece did in the case mentioned above, although problematically, it did so largely in order to gain leverage in its debt negotiations. This may be permissible when the following conditions obtain: (1) the target state is *genuinely overburdened*; (2) it has *clear evidence* that doing so would allow migrants to access protection in other states (and will allow these migrants to return and receive protection if they do not); and (3) doing so is in line with migrants' own mobility preferences. Such a strategy comes with the risk that migrants will not receive access to protection, although the risk is minimized if the above conditions are met.

83 Not all burden-sharing schemes are fair or secure effective protection. Such schemes also interfere with migrants' choices. Although just schemes must take migrants' autonomy interests seriously, migrants' interests in accessing decent international protection takes precedence over their autonomy in choosing a preferred destination. For discussion, see Gibney, "Refugees and Justice Between States."

84 Tsourapas, "The Syrian Refugee Crisis and Foreign Policy Decision-Making in Jordan, Lebanon, and Turkey."

More controversially, it may perhaps be permissible for target states to leverage their ability to allow migrants to transit onwards *in order to recruit funds to benefit these migrants*. Thus, target states may indicate their willingness to displace costs onto their neighbors in order to recruit participation in burden-sharing schemes and additional funding for migrants. This might seem to simply replicate the dilemma with which I began. Is this not just a form of impermissible migration blackmail? Note first that such a policy may be conceptually distinct from migration blackmail as I defined it above, since the aim of such a policy would be to benefit migrants themselves by recruiting funds to host and integrate them rather to use migrants for some independent state objective. When the state acts for these reasons, it does not treat migrants as a mere means. Instead, it leverages migrants' presence in ways designed to promote *their* ends (benefiting them materially and opening up other resettlement options for them). It is thus less clear that this wrongfully instrumentalizes migrants. Similarly, as long as a state engaged in noncooperative bargaining neither forces nor requires migrants to leave but rather offers them protection within their territory, such a policy neither wrongfully threatens migrants nor places them at risk of harm. It is thus not clear that such a policy wrongs migrants. Recall, moreover, that not all cases of leveraging migrants wrong states, and it may be permissible for states who do their fair share of refugee protection to coerce states who fail to do their fair share into sharing responsibility.

While this argument for the permissibility of noncooperative leveraging of migrants is potentially compelling, I am agnostic about whether such actions are fully permissible. They may at best remain morally ambiguous. This is because, on the one hand, these noncooperative strategies may backfire, and on the other hand, these policies may wrongfully contribute to harmful perceptions of migrants as threats. Such tactics thus merit further normative investigation and need to be approached with the utmost caution. However, the larger lesson of this section stands: it is not the case that target states can resist blackmail only by refusing migrants. There are other strategies available to target states. While these strategies may not prove wholly effective and are unlikely to allow target states to avoid being saddled with *some* costs, they may go some way towards enabling states to resist migration blackmail while maintaining a commitment to ensuring that migrants have access to protection.

## 6. Conclusion

I have argued that states faced with migration blackmail sometimes face a difficult policy dilemma: how to resist blackmail while respecting the legitimate interests of migrants. The prevalent response to this dilemma has been for target

states to close their borders and to suspend migrants' rights, including the right to asylum and the standard procedural protections associated with it. I have argued that these measures are generally impermissible for four reasons. First, not all cases of migration blackmail give rise to a dilemma, as allowing migrants leveraged in standard blackmail scenarios to enter and apply for asylum does not usually in fact impose significant burdens on target states or their citizens. Second, target states sometimes bear special responsibilities for the plight of migrants and the circumstances that make blackmail possible. Third, migrants' interests in avoiding harm and accessing protection generally defeat the countervailing interests that target states and citizens have in excluding them. Finally, target states have alternative measures at their disposal to mitigate the costs associated with protecting migrants.

These arguments show that the EU's response to migration blackmail is profoundly mistaken. The EU's approach privileges the interests of target states to the detriment of migrants without adequate justification, problematically framing migrants as a threat. Although proponents of this approach often try to justify it by appealing to the fact that migrants are harmed by instrumentalization, the EU's approach to migration blackmail in fact harm migrants because it displaces the costs of avoiding migration blackmail onto the migrants themselves. Rather than punishing migrants for seeking asylum, the EU should instead open safe migration routes to undercut the leverage of blackmailing states, reverse its policies of externalization, and signal that it does not view migration as a threat by accepting those in need of protection. These are policies that the EU has independent moral and prudential reasons to adopt, and adopting them is required to both prevent and ethically respond to migration blackmail.[85]

*Ludwig-Maximilians-Universität München*
*daniel.sharp@lmu.de*

85 I would like to thank Rainer Bauböck, Julia Mourão Permoser, Martin Ruhs, Lukas Schmid, audiences at the European University Institute and Ludwig-Maximilians-Universität München, and several anonymous reviewers for helpful comments on this paper. I would like to give special thanks to ZEIT-Stiftung's Borders in Focus program, specifically to Anna Hofmann, who organized an informative trip to the Poland-Belarus border. I learned immeasurably from Mateusz Krępa, Maciej Grześkowiak, Kamila Fiałkowska, Magda Kicińska, Karol Grygoruk, Michał Zadara, Kinga Stańcuk, and many other Polish academics, journalists, and activists who are challenging ongoing border violence in Poland.

## REFERENCES


Adamson, Fiona B., and Gerasimos Tsourapas. "Migration Diplomacy in World Politics." *International Studies Perspectives* 20, no. 2 (2019): 113–28.

Aleinikoff, T. Alexander, and David Owen. "Refugee Protection: 'Here' or 'There'?" *Migration Studies* 10, no. 3 (2022): 464–83.

Ancite-Jepifánova, Aleksandra. "Migrant Instrumentalisation: Facts and Fictions." *Verfassungsblog: On Matters Constitutional*, September 21, 2023. https://verfassungsblog.de/migrant-instrumentalisation-facts-and-fictions/.

Balogun, Bolaji. "Refugees Separated by the Global Color Line: The Power of Europeanness, Whiteness, and Sameness." *International Migration Review* (2023).

Bauböck, Rainer, Julia Mourão Permoser, and Martin Ruhs. "The Ethics of Migration Policy Dilemmas." *Migration Studies* 10, no. 3 (2022): 427–41.

Bender, Felix. "Against 'Weaponised Migration'." *Social Europe*, November 10, 2021. https://socialeurope.eu/against-weaponised-migration.

———. "What's Political About Political Refugeehood? A Normative Reappraisal." *Ethics and International Affairs* 36, no. 3 (2022): 353–75.

Betts, Alexander. *The Wealth of Refugees: How Displaced People Can Build Economies*. Oxford University Press, 2021.

Border Violence Monitoring Network. *The Black Book of Pushbacks*. European Parliamentary Group, 2020.

Buxton, Rebecca. "The Duty to Naturalise Refugees." *Critical Review of International Social and Political Philosophy* 26, no. 7 (2023): 1119–39.

Buxton, Rebecca, and Jamie Draper. "Refugees, Membership, and State System Legitimacy." *Ethics and Global Politics* 15, no. 4 (2022): 113–30.

European Council on Refugees and Exiles (ECRE). "ECRE Reaction: No Majority for Instrumentalisation Regulation." *ECRE Weekly*, December 8, 2022. https://ecre.org/ecre-reaction-no-majority-for-instrumentalisation-regulation/.

FitzGerald, David. *Refuge Beyond Reach: How Rich Democracies Repel Asylum Seekers*. Oxford University Press, 2019.

Ganty, Sarah, Aleksandra Ancite-Jepifánova, and Dimitry V. Kochenov. "EU Lawlessness Law at the EU-Belarusian Border: Torture and Dehumanisation Excused by 'Instrumentalisation.'" *Hague Journal on the Rule of Law* 16, no. 3 (2024): 739–74.

Gibney, Matthew J. "Refugees and Justice Between States." *European Journal of Political Theory* 14, no. 4 (2015): 448–63.

Gowayed, Heba. *Refuge: How the State Shapes Human Potential*. Princeton University Press, 2022.

Greenhill, Kelly M. *Weapons of Mass Migration: Forced Displacement, Coercion,*

*and Foreign Policy*. Cornell University Press, 2010.
Grześkowiak, Maciej. "The 'Guardian of the Treaties' Is No More? The European Commission and the 2021 Humanitarian Crisis on Poland-Belarus Border." *Refugee Survey Quarterly* 41, no. 1 (2023): 81–102.
Halemba, Agnieszka. "Europe in the Woods: Reflections on the Situation at the Polish-Belarusian Border." *Ethnologia Europaea* 52, no. 1 (2022): 1–13.
Hayden, Sally. *My Fourth Time, We Drowned: Seeking Refuge on the World's Deadliest Migration Route*. Melville House, 2022.
Herrmann, Therese. "Cosmopolitanism, Global Justice and Refugees." In *Cosmopolitan Norms and European Values: Ethical Perspectives on Europe's Refugee Policy*, edited by Marie Göbel and Andreas Niederberger. Routledge, 2024.
Hidalgo, Javier. "Associative Duties and Immigration." *Journal of Moral Philosophy* 10, no. 6 (2013): 697–722.
Hillier-Smith, Bradley. "Doing and Allowing Harm to Refugees." *Journal of Ethics and Social Philosophy* 18, no. 3 (2020): 298–321.
———. *The Ethics of State Responses to Refugees*. Routledge, 2024.
Howard, Jeffrey W. "Kidnapped: The Ethics of Paying Ransoms." *Journal of Applied Philosophy* 35, no. 4 (2018): 675–88.
Huysmans, Jef. "The European Union and the Securitization of Migration." *Journal of Common Market Studies* 38, no. 5 (2000): 751–77.
Jakob, Christian, and Simone Schlindwein. *Dictators as Gatekeepers for Europe: Outsourcing* EU *Border Controls to Africa*. Translated by Lydia Baldwin and Emal Ghamsharick. Daraja Press, 2019.
Kapelner, Zsolt. "Anti-Immigrant Backlash: The Democratic Dilemma for Immigration Policy." *Comparative Migration Studies* 12, no. 12 (2024).
Kazharski, Aliaksei. "An Authoritarian Spectacle: Visual Biopolitics and the Dramaturgy of the Poland-Belarus Border Migration Crisis." *Visual Anthropology* 36, no. 4 (2023): 373–96.
Kleist, J. Olaf. "Beyond the Crisis Mode of the EU-Turkey Refugee Agreement: Towards Sustainable Rights-Based EU Asylum and Refugee Policies." Policy paper, Heinrich-Böll-Stiftung, March 2022.
Kolodny, Niko. "What Makes Threats Wrong?" *Analytic Philosophy* 58, no. 2 (2017): 87–118.
Koros, Dimitris. "The Normalization of Pushbacks in Greece: Biopolitics and Racist State Crime." *State Crime Journal* 10, no. 2 (2021): 238–56.
Laube, Lena. "Diplomatic Side-Effects of the EU's Externalization of Border Control and the Emerging Role of 'Transit States' in Migration Diplomacy." *Historical Social Research* 46, no. 3 (2021): 78–105.
Lister, Mathew. "Who Are Refugees?" *Law and Philosophy* 32, no. 5 (2013):

645–71.
Majcher, Izabella. "Creeping Crimmigration in CEAS Reform: Detention of Asylum-Seekers and Restrictions on Their Movement under EU Law." *Refugee Survey Quarterly* 40, no. 1 (2021): 82–105.
Micinski, Nicholas R. "Threats, Deportability and Aid: The Politics of Refugee Rentier States and Regional Stability." *Security Dialogue* 54, no. 6 (2023): 529–47.
Mieth, Corinna, and Garrath David Williams. "Beyond (Non)-Instrumentalization: Migration and Dignity Within a Kantian Framework." *Ethical Theory and Moral Practice* 26, no. 6 (2022).
Miller, David. "The Nature and Limits of the Duty of Rescue." J*ournal of Moral Philosophy* 17, no. 3(2020): 320–41.
———. "Responsibility and the Duty of Rescue." *Journal of Applied Philosophy* 39, no. 2 (2022): 313–26.
Muftuler-Bac, Meltem. "Turkey and the European Union Refugee Deal: Assessing Turkish Migration Policies and the External Protection of European Borders." Working Paper, Migration Governance and Asylum Crises (MAGYC) Project, 2020.
Oberman, Kieran. "Is There a Human Right to Immigrate?" In *Migration in Political Theory: The Ethics of Movement and Membership*, edited by Sarah Fine and Leah Ypi. Oxford University Press, 2016.
Owen, David. "Differentiating Refugees: Asylum, Sanctuary, Refuge." In *The Political Philosophy of Refuge*, edited by David Miller and Christine Straehle. Cambridge University Press, 2019.
———. "*In Loco Civitatis*: On the Normative Basis of the Institution of Refugeehood and Responsibilities for Refugees." In *Migration in Political Theory: The Ethics of Movement and Membership*, edited by Sarah Fine and Leah Ypi. Oxford University Press, 2016.
———. "Refugees, Fairness and Taking up the Slack: On Justice and the International Refugee Regime." *Moral Philosophy and Politics* 3, no. 2 (2016): 141–64.
———. *What Do We Owe to Refugees?* Polity Press, 2020.
Parekh, Serene. *No Refuge*: *Ethics and the Global Refugee Crisis.* Oxford University Press, 2020.
Perkowska, M. "(No) Children's Rights at the Polish-Belarusian Border." In *Human Rights and Legal Services for Children and Youth: Global Perspectives*, edited by Asha Bajpai, David W. Tushaus, and Mandava Rama Krishna Prasad. Springer, 2023.
Pettit, Philip. "A Republican Law of Peoples." *European Journal of Political Theory* 9, no. 1 (2010): 70–94.

Ronzoni, Miriam. "The Global Order: A Case of Background Injustice? A Practice-Dependent Account." *Philosophy and Public Affairs* 37, no. 3 (2009): 229–56.

Scanlon, T. M. *Moral Dimensions: Meaning, Permissibility, Blame*. Harvard University Press, 2010.

Schmid, Lukas. "Saving Migrants' Basic Human Rights from Sovereign Rule." *American Political Science Review* 116, no. 3 (2022): 954–67.

Shacknove, Andrew E. "Who Is a Refugee?" *Ethics* 95, no. 2 (1985): 274–84.

Sharp, Daniel. "Collective Self-Determination and Externalized Border Control." *Global Justice: Theory, Practice, Rhetoric* (2025).

———. "Immigration and State System Legitimacy." *Critical Review of International Social and Political Philosophy* 27, no. 2 (2024): 294–304.

———. "Relational Equality and Immigration." *Ethics* 132, no. 3 (2022): 644–79.

———. "The Right to Emigrate: Ethics and Equality in a World of States." *Journal of Ethics and Social Philosophy* 24, no. 3 (2023): 371–408.

Souter, James. *Asylum as Reparation: Refuge and Responsibility for the Harms of Displacement*. Palgrave Macmillan, 2022.

Tittel-Mosser, Fanny. "Reversed Conditionality in EU External Migration Policy: The Case of Morocco." *Journal of Contemporary European Research* 14, no. 4 (2018): 349–63.

Tsourapas, Gerasimos. "The Syrian Refugee Crisis and Foreign Policy Decision-Making in Jordan, Lebanon, and Turkey." *Journal of Global Security Studies* 4, no. 4 (2019): 464–81.

Tsourapas, Gerasimos, and Sotirios Zartaloudis. "Leveraging the European Refugee Crisis: Forced Displacement and Bargaining in Greece's Bailout Negotiations." *Journal of Common Market Studies* 60, no. 2 (2022): 245–63.

Vaagland, Karin, and Oskar Chmiel. "Parochialism and Non-Cooperation: The Case of Poland's Opposition to EU Migration Policy." *Journal of Common Market Studies* 62, no. 5 (2024): 1332–50.

Vinocur, Nicholas, Clea Caulcutt, and Eddy Wax. "Poland Wins After EU Backs Its Proposed Asylum Ban for Russia, Belarus." *Politico*, October 17, 2024. https://www.politico.eu/article/poland-prime-minister-donald-tusk-eu-asylum-ban-russia-belarus-migration-security/.

*Journal of Ethics and Social Philosophy*
VOL. 30, NO. 2 · APRIL 2025
*https://doi.org/10.26556/jesp.v30i2.4053*
© 2025 Author

# RATIONAL INCOMPETENCE OF VOTERS

*Susumu Cato*

THE ESSENTIAL feature of democracy is that all eligible citizens in a society have a right to participate in the collective decision-making process.[1] The efficacy of democracy therefore depends on how citizens vote (and otherwise participate) in this process. Presumably (or ideally, at least), ordinary citizens vote for the candidate or policy they consider best on the basis of social circumstances and information gathered by themselves, but the question is how reliable they (collectively) are.

Epistemic democracy is a philosophical perspective that justifies democracy on the grounds of its ability to make good/correct decisions.[2] The Condorcet jury theorem (CJT) is the central mathematical theorem for epistemic democracy.[3] According to the CJT, decision-making by majority selects the right outcome/policy if citizens decide their votes independently, if each voter is competent to a certain degree, and if the number of citizens involved in the decision-making process is large enough. This theorem has been applied in various ways and contexts and lends support to the idea of epistemic democracy.[4] However, the CJT has been criticized as well. Criticisms are linked to the formal conditions assumed by the CJT. A classical criticism puts the *independence condition*, which requires that the votes of citizens are independent, in doubt. If there is something that causes a strong correlation between votes, the CJT fails.[5] A significant number of studies have considered how such correlation can occur.[6]

1 It is common that individuals must be above a certain age in order to be eligible to vote. In some countries, people convicted of certain crimes may be prohibited from voting.

2 Anderson, "The Epistemology of Democracy," 10–15.

3 For important early work on the CJT, see Black, *The Theory of Committees and Elections*, 159–78; Grofman, "A Comment on 'Democratic Theory,'" 100; and List and Goodin, "Epistemic Democracy," 285.

4 Goodin and Spiekerman, *An Epistemic Theory of Democracy*.

5 The CJT holds under some types/levels of correlation. Berg, "Condorcet's Jury Theorem, Dependency Among Jurors," 91–92; Ladha, "Condorcet's Jury Theorem in Light of de Finetti's Theorem," 77–82; and Pivato, "Epistemic Democracy with Correlated Voters," 59–61.

6 See, for example, Boland, "Majority Systems and the Condorcet Jury Theorem," 185–86; Boland et al., "Modelling Dependence in Simple and Indirect Majority Systems," 83–86;

A relatively new and quite radical criticism concerns the *competence condition*, according to which each voter votes for a right option with a certain probability $p$ that is greater than half. The competence condition assumes that each voter's decision is better than random. In regard to this condition, some critics have argued that citizens are unreasonable, ignorant, and therefore incompetent (i.e., worse than random). For example, Brennan builds his case for epistocracy (decision-making by experts) on this incompetency.[7]

In this paper, I examine the possibility of incompetency among reasonable and non-ignorant agents.[8] I characterize the properties of such agents using the framework of Bayesian rationality. That is, I assume that each citizen updates their belief(s) based on Bayesian inference after getting new, *nonmisleading* information and that each citizen maximizes their expected utility. I show that even with this assumption, there are realistic cases in which agents are incompetent (i.e., the competency condition is violated). Notably, this is a new possibility of voter incompetence, given that unreasonableness or ignorance are the source(s) of voter incompetence in existing work. Importantly, Bayesian voters are not subject to cognitive bias; the possibility of misleading information is excluded.[9] My results demonstrate that the *asymmetry of signals* is a threat to democracy. That said, the asymmetry of signals also disrupts decision-making by epistocrats; thus, it can also be a threat to epistocracy. Moreover, I offer a single example to illustrate that epistocracy can be (conditionally) worse than democracy, based on the asymmetry of signals.

---

Cato and Inoue, "Are Good Leaders Truly Good?" 441–42; and Estlund, "Opinion Leaders, Independence, and Condorcet's Jury Theorem," 138–39.

7 Brennan, *Against Democracy*.

8 Bayesian rational agents satisfy the features of Brennan's "vulcans," who are the ideal types of experts. Brennan writes:

> Vulcans are perfectly rational. An ignorant vulcan would know they are ignorant, and thus would be almost entirely agnostic about political issues. If they decided to learn more, they would seek out information from credible sources. They would conform their beliefs to the best available evidence. A vulcan would look not merely at evidence in favor of different views but also evidence against these views. They would change their minds whenever the evidence called for it. They would consult peers and take disagreement seriously, and would gladly accept criticism, since they want to avoid error. "Thanks for correcting me and pointing out my mistakes!" They would hold beliefs only as strongly as the evidence allows. (*Against Democracy*, 36–37)

9 Brennan mentions two ideal types of incompetent voters: "hobbits" and "hooligans" (*Against Democracy*, 1–22). The former includes voters who do not care about social issues, while the latter includes voters with a strong cognitive bias. Bayesian voters are neither of the two.

This discussion note consists of two main sections. The first shows that Bayesian voters are competent if signals are symmetric; therefore, the CJT works. The second section shows that if signals are asymmetric in a certain way, the same voters can be incompetent, and thus, the CJT fails.

## 1. COMPETENCE OF BAYESIAN AGENTS UNDER SYMMETRIC SIGNALS

Let us assume that there are two states: $L$ and $R$. Each state occurs with a probability 0.5, but voting citizens do not know which state they are actually in. Voters get some signal that represents a relevant piece of information, and there are two such signals, $A$ and $B$. Voters independently receive one of these signals as follows:

- If the true state is $L$, each voter gets signal $A$ with probability $p$ and signal $B$ with probability $1 - p$;
- If the true state is $R$, each voter gets signal $A$ with probability $1 - p$ and signal $B$ with probability $p$.

We assume that $p$ is not equal to 0.5.[10] The probability distribution is shown in table 1. Each voter is assumed to vote for either of the two policies $L$ or $R$. The policy that receives the largest number of votes is collectively selected. If the selected policy coincides with the true state (that is, the right policy is selected), then each voter obtains one unit of utility. Otherwise, they obtain zero. We assume sincere voting and that voters do not have any predetermined ideological bias, so each voter chooses a policy ($L$ or $R$) to maximize their expected utility.[11] I assume that this prior probability structure is commonly known among agents; moreover, the prior distribution is assumed to be correct, and there is no possibility of misleading information.

*Table 1. Symmetric Probability Structure*

| | State $L$ | State $R$ |
|---|---|---|
| Signal $A$ | $p$ | $1 - p$ |
| Signal $B$ | $1 - p$ | $p$ |

10 If this assumption is not satisfied (i.e., $p = 0.5$), then the signals provide no information, which would be essentially the same as no signal.

11 This means that we ignore the possibility of strategic voting, which is examined by Austen-Smith and Banks, "Information, Aggregation, Rationality, and the Condorcet Jury Theorem," 34–35.

It is assumed that each agent knows the probability structure of signals. In a situation like this, voters who apply Bayesian inference update their beliefs about states (and thus about the right policies) after receiving their signals. If an agent obtains $A$ as a signal, they update their belief by applying Bayes's formula:

$$P(L \mid A) = \frac{P(A \mid L)P(L)}{P(A)}$$

$$= p.$$

Analogically, $P(R \mid A) = 1 - p$. This means that if a voter gets signal $A$, then their subjective belief that $L$ is the true state becomes $p$ and that $R$ is the true state becomes $1 - p$. Note that the expected utility of voting for $L$ is $p$, while that of voting for $R$ is $1 - p$. Therefore, if $p$ is larger than 0.5 (resp. $p$ is smaller than 0.5), then voters who receive $A$ vote for $L$ (resp. $R$).

The case with signal $B$ is derived as follows:

$$P(L \mid B) = 1 - p \text{ and } P(R \mid B) = p.$$

If $p$ is larger than 0.5 (resp. $p$ is smaller than 0.5), then voters who receive $B$ vote for $R$ (resp. $L$).

Note that each agent is competent in the sense that they vote for the right policy with a probability that is higher than 0.5 in each state. Now assume that $p$ is greater than 0.5. If $L$ is the actual state, then the probability of getting $A$ as a signal is higher than 0.5. Because a voter votes for $L$ when $A$ is received, the probability that they are correct is higher than 0.5 from the *ex ante* viewpoint. If the actual state is $R$, then the probability of getting $B$ as a signal is larger than 0.5, and again, the probability that they are correct is higher than 0.5. In summary, voters are competent in either state. This is true for any value of $p$ as long as $p$ is not 0.5. To see this, it suffices to consider the case where $p$ is smaller than 0.5. In state $L$, the probability of getting $B$ is higher than 0.5. Then, the voter votes for $L$ when $B$ is received, and the probability that they are correct is higher than 0.5. An analogical argument holds in state $R$. Each voter is competent when $p$ is smaller than 0.5.

## 2. RATIONAL INCOMPETENCE AND ASYMMETRIC SIGNALS

Extending the argument from the previous section, I show in this section how a Bayesian rational agent can be incompetent (even in the absence of misleading information). Significantly, the model in the previous section assumes that the probability distribution for signals is symmetric. Specifically, the probability

of getting $A$ under $L$ is equal to that of $B$ under $R$. (See table 1 again.) The voter competence shown in the previous section depends on this feature.

Now let us consider the following (potentially) asymmetric probability distribution.

- If the true state is $L$, each agent gets signal $A$ with probability $p$ and signal $B$ with probability $1 - p$;
- If the true state is $R$, each agent gets signal $A$ with probability $1 - q$ and signal $B$ with probability $q$.

The probability distribution is shown in table 2. Note that this table coincides with table 1 if $q$ is equal to $p$. However, we assume that $q$ is not equal to $p$.

*Table 2. General Structure of Probability Distributions*

| | State $L$ | State $R$ |
|---|---|---|
| Signal $A$ | $p$ | $1 - q$ |
| Signal $B$ | $1 - p$ | $q$ |

If an agent receives signal $A$, they update their belief in the following manner:

$$P(L \mid A) = \frac{P(A \mid L)P(L)}{P(A)}$$

$$= \frac{p}{p + (1 - q)}.$$

Analogically, $P(R \mid A) = (1 - q)/(p + (1 - q))$. Note that the expected utility of voting for $L$ is $p/(p + (1 - q))$, while that of voting for $R$ is $(1 - q)/(p + (1 - q))$. Therefore, if $p$ is larger than $1 - q$ (resp. $1 - q$ is larger than $p$), then voters who receive $A$ vote for $L$ (resp. $R$).

The case with signal $B$ is derived as follows:

$$P(L \mid B) = \frac{1 - p}{(1 - p) + q} \text{ and } P(R \mid B) = \frac{q}{(1 - p) + q}.$$

If $1 - p$ is larger than $q$ (resp. $q$ is larger than $1 - p$), then voters who receive $B$ vote for $L$ (resp. $R$).

Consider the probability distribution in table 3. According to the voting patterns described above, an agent votes for $L$ if they receive $A$, while they vote for $R$ if they get signal $B$. In state $L$, agents cannot be competent on average even though they update their belief based on Bayesian inference. Notably, each

agent receives signal *A* with probability 0.3 and *B* with probability 0.7. If a voter receives *A*, then they vote for *L*, which is correct. However, if they receive *B*, then they vote for *R*, which is not correct. This means that the voters are correct with probability 0.3; they vote for the wrong policy with probability 0.7. Hence, the voters are incompetent in state *L*. (Note that voters are correct with probability 0.9 in state *R*.) The CJT fails because of this incompetence.

*Table 3. Asymmetric Probability Distributions: Rational Incompetence*

| | State *L* | State *R* |
|---|---|---|
| Signal *A* | 0.3 | 0.1 |
| Signal *B* | 0.7 | 0.9 |

To see that the CJT fails, assume that the number of voters is very large. The share of voters who vote for *L* approximates 0.3 (resp. 0.1) in state *L* (resp. *R*). The share of voters who vote for *R* approximates 0.7 (resp. 0.9) in state *L* (resp. *R*). Hence, it is effectively guaranteed that the wrong policy is selected in state *L*. (On the other hand, it is similarly certain that the correct policy is selected in state *R*.)

It must be emphasized that all of these voters are rational because they follow Bayesian reasoning and maximize their expected utility based on their updated beliefs. Despite this rationality, the probability that a voter's choice is right can be lower than 0.5 in some state. Certainly, Bayesian rational agents can easily hold wrong beliefs when there is sufficient misleading evidence. However, in this case, rational agents tend to fail even if there is neither misperception nor misleading information involved. We can call this phenomenon *rational incompetence*. The cause of this rational incompetence is the probability structure of signals. To be precise, the information each person gets is asymmetric toward *B* (i.e., the probability of getting *B* is higher than 0.5 in both states). This shows that if information in society is unbalanced in a certain way, the competence of votes is compromised even if they are all rational, which can be a threat to (epistemic) democracy.

Notice that this rational incompetence is completely different from rational ignorance, which is well known in the literature on voting. A voter is rationally ignorant when they do not (sufficiently) inform themselves about a political issue because of the cost of obtaining information. In the case of rational incompetence, on the other hand, voters are eager to obtain relevant information and accurately process the information they get. Hence, voters are far from ignorant in such cases.

It is important to understand that asymmetric signals (as in table 3) are not a mere theoretical possibility. On the contrary, such signals are not especially extreme and can be quite realistic. To illustrate, consider a case in which someone who has a fever wants to know whether they have COVID-19. This person will use their symptoms as signals. Here, the prior probability of COVID-19 is assumed to be 50 percent. Assume that 10 percent of people who are infected with COVID-19 experience a new loss of taste or smell; this symptom occurs in less than 1 percent of fever cases that are not caused by COVID-19. If someone experiences a loss of taste or smell, they will believe that there is a high chance of having COVID-19. While this updating is completely rational, the signal structure underlying it is strongly asymmetric, as shown in table 4.

*Table 4. COVID-19 and Its Symptoms*

| | COVID-19 | Not COVID-19 |
|---|---|---|
| Loss of taste or smell | 0.10 | 0.01 |
| No loss | 0.90 | 0.99 |

As another example, consider policymaking associated with nuclear power plants. Nuclear power plants should not be built in areas where there is a high risk of earthquakes. Even in a high-risk area, the probability of the occurrence of a very strong earthquake within the next fifty years is extremely low. (The probability is even smaller outside the high-risk area.) When past earthquake frequency is used as a signal, signal structures are strongly asymmetric. Although earthquakes are not independent signals for voters, it is not unrealistic to assume that a similar asymmetry exists in the case of independent signals such as those considered in the CJT.

Furthermore, the theoretical mechanism for rational incompetence also works if the number of signals is smaller than the number of states. Assume, for example, that there are only two signals, *A* and *B*, but three states, *L*, *M*, and *R*, such that each state occurs with probability 1/3. The probability distribution of the two signals in the three states is shown in table 5. Under this probability distribution, each agent votes for *L* if they receive signal *A*, while they vote for

*Table 5. Three States and Two Signals: Rational Incompetence Due to Signal Insufficiency*

| | State *L* | State *M* | State *R* |
|---|---|---|---|
| Signal *A* | 0.7 | 0.6 | 0.3 |
| Signal *B* | 0.3 | 0.4 | 0.7 |

*R* if they get *B*. Hence, no one votes for *M*, even when *M* is the actual state. This implies that all voters will be completely incompetent in *M*; therefore, the CJT fails. Arguably, it is likely that the problem of asymmetry becomes more severe when there are more states.

Lastly, the same rational incompetence can also affect epistocracy. In *restricted suffrage*, which has been suggested by advocates of epistocracy, only a finite number of qualified experts will vote.[12] Presumably, these experts will be rational and will use the knowledge they have gained to decide on the options they can vote for. However, if there are asymmetries in the signals, then there is a good chance that their rational voting will not increase the probability that the right policy is chosen. Hence, epistocracy may not work as well as assumed by its advocates. Notably, in case of asymmetric signals, the rationality of agents does not help to increase the probability of choosing the right policy; thus, restricting voting rights on grounds of irrationality is not a plausible method to assure the best outcome.

Next let us examine a scenario in which epistocracy does not have a strict competency advantage over democracy. In the situation described in table 3, I assume that experts vote based on Bayesian updating, while nonexperts vote randomly without any deliberate thought. In other words, experts vote for *L* when they receive signal *A* and for *R* when they receive signal *B*. By contrast, nonexperts vote for either option with equal probability.[13] In this scenario, the CJT fails for both democracy and epistocracy, but the failure manifests differently in each regime. In the case of epistocracy, the majority believes that *R* is likely to be true, regardless of the true state; therefore, an incorrect outcome tends to be selected when the true state is *L*. By contrast, under democracy, where nonexperts vote randomly, each outcome is selected with probability 0.5, regardless of the true state. Consequently, when the true state is *L*, democracy is more likely to yield the correct outcome compared with epistocracy. Thus, one could argue that democracy is *conditionally* better than epistocracy when the true state is *L*.[14]

## 3. CONCLUDING REMARKS

This discussion note has shown that people can easily become incompetent voters depending on the probability distribution of signals (i.e., information).

12 See Brennan, *Against Democracy*, 213–18.

13 This assumption aligns with Brennan's concept of "hobbits." See note 9 above.

14 Notably, this does not imply that epistocracy is unconditionally worse than democracy. Both regimes have an unconditional (or *ex ante*) probability of 0.5 for making the correct choice.

This paper's contribution is not only that it demonstrates the possibility that the competency condition is not satisfied but also that it identifies how it is not satisfied. Importantly, my analysis is not based on a simple observation that Bayesian rational agents believe the wrong information if they receive sufficient misleading evidence. I demonstrate that the competency condition may fail even if an agent gets the right information (i.e., that a loss of taste is a sign of COVID-19). This situation is likely to occur when signal availability is limited. Hence, what the CJT actually requires is that sufficient informational signals are available for all options and possibilities.

The information structure identified in the paper is indeed likely to occur in real-world scenarios. For instance, consider a scenario in which a candidate might have dementia. Even if this were the case, the likelihood of voters receiving a clear signal indicating dementia is low. Generally, signals suggesting a candidate's unsuitability or highlighting fundamental policy errors are often not easily accessible, even to experts. This leads to an asymmetric information structure. Such asymmetries can result in voters making wrong decisions, thereby significantly impacting the democratic process. It is crucial to understand and acknowledge the prevalence of these information asymmetries to ensure a nuanced debate on the merits and challenges of both epistocratic and democratic systems.[15]

*University of Tokyo*
*susumu.cato@gmail.com*

## REFERENCES

Anderson, Elizabeth. "The Epistemology of Democracy." *Episteme* 3, nos. 1–2 (2006): 8–22.

Austen-Smith, David, and Jeffrey S. Banks. "Information, Aggregation, Rationality, and the Condorcet Jury Theorem." *American Political Science Review* 90, no. 1 (1996): 34–45.

Berg, Sven. "Condorcet's Jury Theorem, Dependency Among Jurors." *Social Choice and Welfare* 10, no. 1 (1993): 87–95.

Black, Duncan. *The Theory of Committees and Elections*. Cambridge University Press, 1958.

Boland, Phillip J. "Majority Systems and the Condorcet Jury Theorem." *Jour-*

15 I would like to thank Lajos Brons, Akira Inoue, Shu Ishida, and Ken Oshitani for their helpful comments. I gratefully acknowledge KAKENHI for its financial support (grant nos. JP18H05204, JP20H01446, JP22K01387, JP22H05083, JP22H05086, AND JP23K01316).

*nal of the Royal Statistical Society Series D: The Statistician* 38, no. 3 (1989): 181–89.

Boland, Phillip J., Frank Proschan, and Yung Liang Tong. "Modelling Dependence in Simple and Indirect Majority Systems." *Journal of Applied Probability* 26, no. 1 (1989): 81–88.

Brennan, Jason. *Against Democracy*. Princeton University Press, 2016.

Cato, Susumu, and Akira Inoue. "Are Good Leaders Truly Good?" *Analysis* 83, no. 3 (2023): 437–46.

Estlund, David M. "Opinion Leaders, Independence, and Condorcet's Jury Theorem." *Theory and Decision* 36 (1994): 131–62.

Goodin, Robert E., and Kai Spiekermann. *An Epistemic Theory of Democracy*. Oxford University Press, 2018.

Grofman, Bernard. "A Comment on 'Democratic Theory: A Preliminary Mathematical Model.'" *Public Choice* 21, no. 1 (1975): 99–103.

Ladha, Krishna K. "Condorcet's Jury Theorem in Light of de Finetti's Theorem: Majority-Rule Voting with Correlated Votes." *Social Choice and Welfare* 10, no. 1 (1993): 69–85.

List, Christian, and Robert E. Goodin. "Epistemic Democracy: Generalizing the Condorcet Jury Theorem." *Journal of Political Philosophy* 9, no. 3 (2001): 277–306.

Pivato, Marcus. "Epistemic Democracy with Correlated Voters." *Journal of Mathematical Economics* 72 (2017): 51–69.

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